

Der C-Test als Lesetest bei Muttersprachlern

Dissertation

zur Erlangung des akademischen Grades

Doktorin der Philosophie (Dr. Phil.)

Vorgelegt dem Fachbereich

Geisteswissenschaften

an der

Universität Duisburg-Essen

von

Anna Mashkovskaya, M.A.

geb. am 05.12.1980 in Tomsk

1. Gutachter: 1. Prof. Dr. Rupprecht S. Baur, Universität Duisburg-Essen
2. Gutachter: 2. Prof. Dr. Albert Bremerich-Vos, Universität Duisburg-Essen

Disputation am 4.12.2013

Essen, Juli 2013

Inhaltsverzeichnis	I
Einleitung	1
1. Sprachkompetenz und Sprachkompetenzmodelle	6
1.1 Was ist Kompetenz?	6
1.2 Grundmerkmale der Kompetenz	8
1.3 Zum Begriff Sprachkompetenz	11
1.4 Entwicklung des Sprachkompetenzbegriffs	13
1.5 Modelle zur Erfassung von Sprachkompetenz	15
1.6 Vorstellung ausgewählter Studien zur Erfassung von Sprachkompetenz	19
1.6.1 PISA-Studie	20
1.6.2 DESI-Studie	25
2. Lesekompetenz und Lesemodelle	33
2.1 Was ist Lesekompetenz	33
2.2 Lesemodelle	37
2.2.1 Lesemodell von Artelt et al. (2007)	37
2.2.2 Lesemodell von Rosebrock & Nix (2008)	44
2.2.3 Lesemodell von Richter & Christmann (2002)	48
2.2.4 Verortung von Tests in Lesemodellen	51
3. Sprachtests	54
3.1 Ansätze zur Konstruktion von Sprachtests	54
3.2 Sprachtests und ihre Anwendungsbereiche	58
3.3 Sprachtests in der Schule	62
3.4 Sprachtests für erwachsene Sprecher an der Universität	72
3.5 Sprachtests für den Hochschulzugang ausländischer Studienbewerber	75
4. C-Tests	81
4.1 Kommunikationstheorie und Redundanz	81
4.2 Übertragung der Redundanz auf die Testtheorie	83
4.3 Übertragung der Redundanz auf die Formate Cloze-Test und C-Test	84
4.3.1 Cloze-Tests	84
4.3.2 C-Tests	87
4.4 Entwicklung eines klassischen C-Tests	88
4.5 Modifizierte C-Tests	91
4.6 Einsatzmöglichkeiten von C-Tests und Gründe für den Einsatz im akademischen Bereich	94

4.7 Kritik an C-Tests	103
4.8 Überprüfung von Testgütekriterien von C-Tests für Studierende	112
4.9 Auswertung und Interpretation von C-Test-Ergebnissen	115
5. Lesetests	120
5.1 Lesetestformate und grundlegende Fragen vor dem Einsatz von Lesetests	120
5.2 Überblick über Lesetests für Deutsch als Muttersprache (DaM)	127
5.3 Überblick über Lesetests für Deutsch als Fremdsprache (DaF)	133
5.4 Auswahl an Lesetests für die vorliegende Studie	137
5.4.1 Testkombination aus Multiple-Choice-Lesetests	137
5.4.2 Testkombination aus Stolperwörter-Lesetests	147
6. Überlegungen zur Verbindung von C-Tests und Lesetests	155
6.1 Können C-Tests Lesefähigkeiten messen?	156
6.2 Prozessebenen bei der Bearbeitung von C-Tests	161
6.3 Kategorisierung der Lücken	163
6.4 Kategorisierung der Lücken an einem C-Test-Beispiel	165
6.5 Hypothesen zur Differenzierung zwischen den Probanden auf den Prozessebenen	169
7. Datenerhebung und Analysen	176
7.1 C-Tests und Lesetests: Gruppen und Daten	177
7.2 C-Tests: Item- und Reliabilitätsanalysen	179
7.3 Vergleich zwischen den C-Tests und Lesetests: Reliabilitätsanalysen und Korrelationsberechnungen	185
7.3.1 C-Tests und LT1 (schulischer Stolperwörtertest)	185
7.3.2 C-Tests und LT2 (modifizierter Stolperwörtertest für Erwachsene)	188
7.3.3 C-Tests und LT3 (Stolperwörtertest für Erwachsene mit veränderter Reihenfolge)	193
7.3.4 C-Tests und LT4 (MC-Lesetest und MC-Zuordnungsaufgabe)	196
8. Hypothesen: Ergebnisse und Kommentare	201
8.1 Hypothesen zu den Hintergrundvariablen	201
8.1.1 C-Testleistung und Geschlecht	201
8.1.2 C-Testleistung und Migrationshintergrund	205
8.1.3 C-Testleistung und Alter	213
8.1.4 C-Testleistung und Schulabschluss	217
8.1.5 C-Testleistung und angestrebter Schulabschluss	221
8.2 Hypothesen zur Validität von C-Tests	224
8.2.1 C-Testleistung und Deutschnote	225

8.2.2 C-Testleistung und Englische note	229
8.2.3 C-Testleistung und Selbsteinschätzung der Sprachkompetenz	231
8.2.4 C-Testleistung und Selbsteinschätzung der Lesekompetenz	235
8.2.5 C-Testleistung und Selbsteinschätzung der schriftlichen Kompetenz	239
9. Abschließender Überblick	243
Literaturverzeichnis	253

Verzeichnis der Abbildungen im Text

Abb. 1: Modell der vier Grunddimensionen von Sprachkompetenz	16
Abb. 2: Darstellung der eindimensionalen und mehrdimensionalen Modelle zur Sprachkompetenzstruktur	17
Abb. 3: Determinanten der Lesekompetenz nach Artelt et al. (2007)	37
Abb. 4: Mehrebenenmodell von Rosebrock & Nix (2008)	45
Abb. 5: Gesamtmodell DESI-Kompetenzen	70
Abb. 6: Beispiel für die Normierung der Daten	116
Abb. 7: Kategorisierung der Lücken auf der Wortebene	166
Abb. 8: Kategorisierung der Lücken auf der Satzebene	167
Abb. 9: Kategorisierung der Lücken auf der Textebene	168
Abb. 10: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der Wortebene	170
Abb. 11: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der Satzebene	171
Abb. 12: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der Textebene	172
Abb. 13a: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der hierarchieniedrigen Prozessebene	173
Abb. 13b: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der höheren Prozessebene	174
Abb. 14: C-Testergebnisse männlicher und weiblicher Studierender (Gesamtgruppe)	203
Abb. 15: C-Testleistungen der Studierenden mit und ohne Migrationshintergrund	209
Abb. 16: C-Testleistung und erworbener Schulabschluss	218
Abb. 17: C-Testleistung vs. angestrebter Abschluss	222
Abb. 18: C-Testleistung vs. Abiturnote im Fach Deutsch	227
Abb. 19: C-Testleistung vs. Abiturnote im Fach Englisch	230

Abb. 20: C-Testleistung vs. globale Selbsteinschätzung der Sprachkompetenz im Deutschen	232
Abb. 21: Durchschnittliche C-Testleistungen vs. Selbsteinschätzung der Sprachkompetenz im Deutschen	233

Verzeichnis der Tabellen im Text

Tabelle 1: Darstellung der PISA-Dimension (Lesen)	22
Tabelle 2: Bestandteile des DESI-Tests (Fach Deutsch)	26
Tabelle 3: Bestandteile des DESI-Tests (Fach Englisch)	27
Tabelle 4: Überblick über Vorteile und Nachteile von C-Tests	97
Tabelle 5: Beispiel für die Festlegung der Leistungsgruppen	118
Tabelle 6: Lesetestverfahren für die Sprecher des Deutschen als Muttersprache (DaM)	130
Tabelle 7: Lesetestverfahren für die Sprecher des Deutschen als Fremdsprache (DaF)	134
Tabelle 8: Testkombination aus C-Tests und Multiple-Choice-Lesetests	144
Tabelle 9: Analyse der zu messenden sprachlichen Teilkompetenzen mit verschiedenen Testverfahren	145
Tabelle 10: Einsatz des Stolperwörtertests für Erwachsene	149
Tabelle 11: Testset aus C-Tests zuerst und Stolperwörtertest danach	152
Tabelle 12: Testset aus Stolperwörtertest zuerst und C-Tests danach	152
Tabelle 13: Mit den C-Tests und dem Stolperwörtertest getestete sprachliche Teilkompetenzen	153
Tabelle 14: Erhebungen an der Universität Duisburg-Essen	176
Tabelle 15: Das Testdesign der Studie und die Zusammensetzung der Studierenden in verschiedenen Stichproben	179
Tabelle 16: Mittelwerte, Schwierigkeiten und Standardabweichungen von C-Tests	180
Tabelle 17: Interkorrelationen der einzelnen C-Tests	182
Tabelle 18: Festlegung der Daten für die C-Tests für Lehramtsstudierende	184
Tabelle 19: Korrelationen zwischen den C-Tests und dem Stolperwörtertest (LT1)	187
Tabelle 20: Korrelationen zwischen den C-Tests und dem Stolperwörtertest (LT2)	192

Tabelle 21: Korrelationen zwischen den C-Tests und dem Stolperwörtertest (LT3)	194
Tabelle 22: Korrelationen zwischen den C-Tests und den MC-Lesetests	197
Tabelle 23: Anteil von weiblichen und männlichen Probanden auf den Leistungsstufen in den C-Tests (Gesamtgruppe)	205
Tabelle 24: Die größten Sprachgruppen in der Gesamtgruppe	207
Tabelle 25: C-Testleistungen der Studierenden mit und ohne Migrationshintergrund, aufgeteilt nach Leistungsstufen	210
Tabelle 26: C-Testleistungen, aufgeteilt nach Leistungsstufen und Sprachgruppen	211
Tabelle 27: C-Testleistungen in verschiedenen Alterskategorien	214
Tabelle 28: C-Testleistungen der Studierenden zwischen 19 und 25 Jahren	215
Tabelle 29: C-Testleistung der Studierenden aufgeteilt nach Alterskategorien und abgeschlossenem Studium	216
Tabelle 30: C-Testleistung vs. erworbener Schulabschluss	220
Tabelle 31: C-Testleistungen Studierender verschiedener Lehramtsrichtungen aufgeteilt nach Leistungsstufen	223

Einleitung

C-Tests haben sich in der Testforschung als ein zuverlässiges und praktikables Instrument zur Erfassung der allgemeinen Sprachkompetenz etabliert. Bis 2009 lagen keine Studien vor, die den C-Test zur Messung der muttersprachlichen Kompetenz im Deutschen von Studierenden¹ einsetzten. Erstmals erprobte Wockenfuß (2009) C-Tests an erwachsenen Muttersprachsprechern im Rahmen ihrer Dissertation; allerdings wurden C-Tests von ihr nicht zur Diagnostik oder Messung sprachlicher Fähigkeiten, sondern zur Erfassung von Intelligenzleistungen verwendet. In der Studie zeigte sich jedoch, dass C-Tests als Format flexibel sind und auch für erwachsene Muttersprachler adaptiert werden können.

Den Weg der Adaption von C-Tests für erwachsene Muttersprachler und Zweitsprachler mit quasi muttersprachlichen Kompetenzen hat ein Projektteam der Universität Duisburg-Essen weiter verfolgt. Im Rahmen des vom BMBF finanzierten *Sprachkompetenzprojekts Diagnose & Förderung* entstand die im Folgenden vorgestellte Untersuchung. Ihr Ziel ist es, die globalen schriftsprachlichen Kompetenzen von Lehramtsstudierenden zu überprüfen. Ein vollständigeres Bild der Schriftsprachkompetenz der Lehramtsstudierender ergibt sich im Zusammenspiel von zwei Testinstrumenten, die innerhalb des Projekts entwickelt wurden: dem C-Test, der als Screening-Instrument sprachlich leistungsschwächere Studierende identifizieren soll und einer umfangreicheren Schreibaufgabe², durch die sprachliche Fähigkeiten diagnostisch erfasst werden. Im Rahmen des Projektes werden u.a. Korrelationen zwischen der Schreibaufgabe und den C-Tests überprüft. Bei ausreichend hohen Korrelationen zwischen beiden Testaufgaben könnte eine Basis geschaffen werden, um mit dem C-Test die schriftlichsprachliche Leistung von Studierenden vorherzusagen. Sollte dies gelingen, so läge mit dieser Testkombination ein Verfahren vor, mit

¹ Begriffe wie Studierende, Probanden oder Testpersonen werden in dieser Arbeit im generischen Sinne verwendet. Im Weiteren sind unter dem Begriff Schüler (Lehrer) sowohl Personen männlichen als auch weiblichen Geschlechts zusammengefasst.

² Die Schreibaufgabe verlangt von den Studierenden, einen argumentativen Text (hier: einen Zeitungsartikel aus dem Feuilleton einer überregionalen Tageszeitung) zusammenfassen (vgl. Scholten-Akoun, Kuhnen & Mashkovskaya, 2012).

dem auch größere Gruppen von Studierendenanfängern effizient und ohne großen Aufwand getestet werden können. In der Folge könnte den auf diese Weise festgestellten sprachlichen Problemen durch gezielte Fördermaßnahmen begegnet werden.

In der hier vorgelegten Arbeit soll untersucht werden, ob der C-Test als Testformat zur Überprüfung der Lesekompetenz der Studierenden geeignet ist und ob sich mit ihm auch die Lesefähigkeit der Studierenden differenziert erfassen lässt.

Um die Korrelation mit der Lesefertigkeit herzustellen, werden die C-Tests zunächst zusammen mit adressatenspezifischen Lesetests eingesetzt. Überprüft wird, ob die Ergebnisse von C-Tests bei erwachsenen Muttersprachlern mit der Lesefertigkeit korrelieren und ob für die Lösung der C-Test-Lücken die Beherrschung von Lesestrategien auf unterschiedlichen Ebenen nachgewiesen werden kann. Durch Zuordnung der C-Test-Lücken zu verschiedenen Ebenen soll die Beherrschung der Lesestrategien einzelner Lerner erfasst werden. Lassen sich bei der Verknüpfung der Items mit Lesestrategien auf unterschiedlichen Prozessebenen Unterschiede zwischen den Studierenden erkennen, ist zu analysieren, ob die Leistungsdifferenzen zwischen leistungsstarken und leistungsschwachen Probanden mit Lösungsprozessen auf unterschiedlichen Ebenen verbunden sind. Wenn sich mit den C-Tests Defizite in den Lesefähigkeiten der Studierenden diagnostizieren lassen, könnte dies als Indiz dafür gewertet werden, dass neben der Förderung der Schreibfertigkeit auch eine gezielte Förderung unzureichend ausgebildeter Lesestrategien sinnvoll sein könnte.

Während viele empirische Untersuchungen im Bereich Deutsch als Fremdsprache von hohen Korrelationen zwischen dem C-Test und dem Leseverständnistests berichten (vgl. Arras et al., 2002, Eckes & Grotjahn, 2006a,b und ausführlich Kap. 4), gibt es diesbezüglich noch keine Daten für das Deutsche als Muttersprache. In der vorliegenden Untersuchung wird diese Forschungslücke geschlossen, indem dieser Zusammenhang erstmals auch bei erwachsenen Muttersprachlern mit voll entwickelter Sprachfähigkeit untersucht wird. Würden Korrelationen zwischen den C-Tests und den verwendeten Lesetests theoretisch und statistisch

nachgewiesen, so ließen sich C-Test-Ergebnisse als Prädiktoren zur Einschätzung der Lesefähigkeit interpretieren.

Die vorliegende Arbeit gliedert sich in folgende Abschnitte: Im Anschluss an die Einleitung stellt *Kapitel 1* die Entwicklung des Kompetenz- und Sprachkompetenzbegriffs vor. Hierzu gehören grundlegende theoretische Annahmen über die Grundeigenschaften von Kompetenz. Beginnend mit Definitionsansätzen zur Sprachkompetenz allgemein, werden Modelle zur Erfassung der Sprachkompetenz in der Erst- und Fremdsprache dargestellt. Im Anschluss werden zwei große Leistungsstudien in Bezug auf ihre Test- und Messkonzepte zur Erfassung von Sprachkompetenz hin untersucht: PISA und DESI.

In Kapitel 2 wird der Begriff Lesekompetenz dargestellt. Im Fokus der Analyse stehen drei Lesemodelle, welche die beim Lesen sich vollziehenden Prozesse auf unterschiedlichen Ebenen beschreiben. In Anbindung an die Anforderungen der Lesemodelle werden geeignete Sprachtests ausgewählt, die die in den Modellen beschriebenen Prozesse operationalisieren sollen.

Kapitel 3 liefert einen Überblick über die im deutschen Sprachraum zur Verfügung stehenden Sprachtests im schulischen und universitären Bereich, die in Bezug auf die Anforderungen an die Tests, die im Rahmen der vorliegenden Studie zum Einsatz kommen können, analysiert werden.

In *Kapitel 4* wird das Format C-Tests dargestellt, das für die Untersuchung der Lesekompetenz von Lehramtsstudierenden die Grundlage der Analysen bietet. Ausführlich wird auf ihre Entwicklung, Durchführung, Auswertung und Kritik eingegangen. Besonderer Augenmerk richtet sich dabei auf die Überprüfung von Gütekriterien.

In *Kapitel 5* werden die im Bereich Deutsch als Mutter- und Fremdsprache verwendeten Lesetests dargestellt und in Bezug auf ihre Ziele und Grundkonzepte hin analysiert, um eine für die Zielgruppe geeignete Auswahl an Lesetests zu treffen, die zur Überprüfung der Lesefähigkeiten der Studierenden

im Rahmen der vorliegenden Studie zusammen mit den C-Tests eingesetzt werden können.

In *Kapitel 6* wird der Frage nachgegangen, ob C-Tests Lesefähigkeiten abbilden können und falls ja, auf welchen Leseprozessebenen dies geschieht. Mit Hilfe der in Kapitel 2 dargestellten Lesemodelle wird eine Kategorisierung der C-Test-Lücken vorgenommen, um zu zeigen, welche Prozessebenen bei der Rekonstruktion von C-Tests eine Rolle spielen und bei welchen Prozessen leistungsschwächere Studierende Schwierigkeiten haben. Im Anschluss an die Analysen werden die Hypothesen in Bezug auf Leistungsunterschiede beim Lesen überprüft. In Kapitel 6.4 werden die Ergebnisse der Analysen dargestellt, die in Bezug auf die Bestimmung der Validität von C-Tests wichtig sind.

Kapitel 7 geht auf das methodische Vorgehen der vorliegenden Studie ein (Größe der Stichprobe, Einsatz und Durchführung von Tests). Die Hypothesen in Bezug auf die C-Tests und Lesetests werden überprüft und ausführlich erörtert. Im Fokus stehen dabei vor allem Korrelationsberechnungen, die überprüfen, ob die C-Tests und die verwendeten Lesetests gleiche bzw. unterschiedliche Teildimensionen der Lesekompetenz messen. Bestehen entsprechende Korrelationen, kann die Validität von C-Tests für erwachsene Muttersprachler dadurch gesichert werden.

Kapitel 8 stellt eine Reihe von Hypothesen vor, die im Rahmen der vorliegenden Studie anhand der zusätzlich durch Fragebögen erhobenen persönlichen Daten der Studierenden erfasst wurden. Ein Teil der Hypothesen bezieht sich auf die Überprüfung der Konstruktvalidität der verwendeten C-Tests für Studierende und überprüft die Zusammenhänge zwischen den C-Testergebnissen und z. B. Abiturnoten im Fach Deutsch und die Selbsteinschätzung der Kompetenzen. Der zweite Teil der Hypothesen vergleicht die Testergebnisse unter Berücksichtigung unterschiedlicher Gruppen (z. B. Studierende mit und ohne Migrationshintergrund oder mit verschiedenen Schulabschlüssen). Die erhobenen Daten werden in Verbindung mit den C-Testleistungen gesetzt und auf eine Signifikanz möglicher Leistungsunterschiede hin analysiert.

Eine Diskussion in *Kapitel 9* schließt die Arbeit ab mit einer Zusammenfassung relevanter Resultate und ihrer Bedeutung für das *Sprachkompetenzprojekt*

Diagnose & Förderung an der Universität Duisburg-Essen und die C-Test-Forschung allgemein.

1.1 Was ist Kompetenz?

Der Begriff „Kompetenz“ erfährt in Folge gesellschaftlicher Entwicklungen (schulische Anforderungen, schulische und unterrichtliche Qualitätsentwicklungen, Schulleistungsmessungen, z. B. TIMSS, PISA, DESI) heute eine große Bedeutung und Aufmerksamkeit: Kompetenz wird in unserer Gesellschaft als eine wesentliche Voraussetzung verstanden, um im beruflichen und privaten Leben erfolgreich zu handeln.

Der Kompetenzbegriff hat eine lange Geschichte. Experten der Kompetenzforschung sind mittlerweile der Ansicht, dass Kompetenz mehrdimensional ist, weshalb es verfehlt sei, auf ein einheitliches Verständnis von Kompetenz zu hoffen. Sie bezeichnen den Kompetenzbegriff als „theorierelativ“: Kompetenz hat danach jeweils „nur innerhalb der spezifischen Konstruktion einer Theorie von Kompetenz eine definierte Bedeutung“ (Erpenbeck & Rosenstiel 2003, S. 24). Das heißt, dass das Theoriekonstrukt, das der jeweiligen Messung zugrunde liegt, im Wesentlichen die für den Kompetenzbegriff relevanten Bereiche bestimmt. Möchte man z. B. Sprache messen, wird der Kompetenzbegriff ein anderer sein als bei der Messung fachlicher Kompetenzen physikalischer oder musikalischer Art. Im Folgenden wird auf die Definitionen eingegangen, die für die vorliegende Studie von Relevanz sind.

Nach Weinert (2001) wird der Kompetenzbegriff folgendermaßen definiert: Kompetenzen sind

„die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert, 2001, S. 27).

Klieme et al. (2003) beschreiben den Kompetenzbegriff ähnlich; sie betonen seine wichtige Komponente *Motivation des Handelnden*, die den gesamten Handlungsprozess bestimmt und Auswirkungen auf die zu erbringende Leistung besitzt. Kompetenz umfasst danach netzartig zusammenwirkende Facetten wie Wissen, Fähigkeit, Verstehen, Können, Handeln, Erfahrung und Motivation. Sie

1. Sprachkompetenz und Sprachkompetenzmodelle

wird verstanden als Fähigkeit, konkrete Anforderungssituationen zu bewältigen und äußert sich in der erbrachten Leistung. Die Autoren betonen, dass die vorhandenen Kompetenzen das Ergebnis bisheriger Lernprozesse sind. Sie sollen nicht mehr als kurzfristig trainierbar gelten, sondern als Ausdruck einer dynamischen Wissens- und Handlungsbasis verstanden werden, die als Potential vorhanden sind (vgl. Klieme 2003, S. 72 f.).

Carle (2003) entwickelt diese Definition weiter; dabei legt die Autorin großen Wert auf die Trennung zwischen individuell und kollektiv Handelnden. Nach Carle existieren Erfahrungen in Form von Wissen, Können und Einstellungen, die als Ressourcen einer oder mehreren Personen zur Verfügung stehen:

„Kompetenz beschreibt die Handlungsfähigkeit eines individuellen, kollektiven oder organisationalen Akteurs in einer Arbeitssituation und formt sich durch die Entwicklung der Handlungsorganisation in prototypischen Arbeitssituationen - beiläufig oder in einem gezielten Bildungsprozess. Kompetenz leistet die aufgabenorientierte Integration der personalen, kollektiven, organisationalen Erfahrungen des Könnens, Wissens, der Werte, Einstellungen und der Selbstreflexion sowie der situativ verfügbaren Handlungsbedingungen und Mittel zu erfolgsrelevanten Handlungsinterfaces.“ (Carle, 2003, S. 28)

Die oben beschriebenen Definitionen betonen die Handlungsfähigkeit in konkreten Situationen und sehen Kompetenzanwendende vor allem als ‚Akteure‘. Für ein besseres Verständnis der Definition sollen vor der Erfassung von Kompetenzen die Handlungssituationen bestimmt und beschrieben werden. Im Hinblick auf das Handeln kann Kompetenz zusammenfassend als Repertoire verschiedener Ressourcen beschrieben werden: Fertigkeiten, Fähigkeiten, Erfahrungen, physische (z. B. Kraft, Schnelligkeit), psychische (z. B. Stressresistenz), emotionale und soziale Ressourcen: Temperament, Begabungen, Interessen, Motivation etc. Das Individuum kann seine Kompetenzen kombinieren und weiterentwickeln, um in entsprechenden Situationen erfolgreich zu handeln.

1.2 Grundmerkmale der Kompetenz

In der wissenschaftlichen Diskussion haben sich verschiedene Grundmerkmale von Kompetenz durchgesetzt, die insbesondere für die Erfassung und Interpretation von Kompetenz von zentraler Bedeutung sind³.

1. Das Vorhandensein von Kompetenz wird durch das Handeln sichtbar, die Kompetenz ist nicht direkt messbar

Kompetenz ist nicht direkt beobachtbar, sondern lediglich „aus der Realisierung der Dispositionen erschließbar und evaluierbar“ (Erpenbeck & Heyse, 1999, S. 24); sie lässt sich nur durch „ihre Manifestation in Handlungen interpretierend erschließen“ (Baitsch et al., 1998, S. 7). Daher kann Kompetenz nur aus der Handlung heraus erschlossen werden, d. h. sie kann beobachtet, erfasst und bewertet werden. Handeln als Kategorie stellt somit den Ansatzpunkt für die Entwicklung entsprechender Erfassungsverfahren dar. Aus dem genannten Handlungsbezug von Kompetenz ergibt sich für die Erfassung und Bewertung von Kompetenz die Notwendigkeit, sie im Zusammenhang mit dem Handlungskontext zu betrachten (vgl. Schröder 2007, S. 290; Nodari 2002, S. 2).

2. Kompetenz ist situations- und kontextabhängig

Die Kompetenz einer Person wird in Abhängigkeit von den vorgegebenen Anforderungen aktiviert. Kompetenzen, die im Repertoire einer Person vorhanden sind, aber in der entsprechenden Situation nicht benötigt werden, bleiben für Dritte verborgen und sind der Erfassung nicht zugänglich (vgl. Moore & Theunissen, 1994). Daher werden z. B. bei fremdsprachlichen Prüfungen verschiedene Kompetenzbereiche überprüft: Hören, Sprechen, Lesen und Schreiben. Wird ein Kompetenzbereich nicht erfasst, so kann u. a. auf die Kompetenz in diesem Bereich durch die Ausprägung eines anderen Kompetenzbereichs geschlossen werden, da sich die Teilkompetenzen der

³ Eine ausführliche Beschreibung der Kompetenz findet sich bei Kaufhold, 2006, S. 22 ff.

1. Sprachkompetenz und Sprachkompetenzmodelle

Sprachkompetenz in gewisser Abhängigkeit voneinander entwickeln. Ist die Fähigkeit, geschriebene Texte gut zu verstehen und in kontextrelevanten Situationen zu nutzen, vorhanden und gut ausgebildet, so kann davon ausgegangen werden, dass die Schreibfähigkeit der betreffenden Person auf einem vergleichbaren Niveau entwickelt ist. Gleiches gilt für die Teilkompetenzen Hören und Sprechen.

Die Ausprägung einer Kompetenz, d. h. ihr Sichtbarwerden, kann je nach Funktion und Anspruchsniveau der Situation unterschiedlich sein. Die Bestimmung von Kompetenz ist nur in solchen Situationen möglich, in denen sie auch tatsächlich gefordert ist und zur Anwendung kommt.

In Abhängigkeit vom Verwendungskontext wird das geeignete Testformat gewählt, mit dem das ausgewählte Merkmal getestet werden kann. Diese zwei Kriterien (Situation und Kontext) werden bei der Erfassung von Kompetenzen besonders hervorgehoben, da sie durch Auswahl geeigneter (u. a. ‚authentischer‘) Aufgaben zum einen die Generalisierung im Hinblick auf die Fähigkeiten zur Lösung analoger Probleme außerhalb der Testsituation ermöglicht und zum anderen die Akzeptanz der Testaufgaben durch die Probanden erhöht (vgl. Grotjahn, 2003, S. 28).

Für die Entwicklung von Kompetenzerfassungsverfahren ist es notwendig, die jeweilige (Handlungs-)Situation exakt zu beschreiben, um die zu erfassende Kompetenz in den Kontext einzuordnen und dementsprechend zu betrachten und einzuschätzen (vgl. Kaufhold, 2006, S. 24).

3. Kompetenz ist kein statisches Konstrukt, sie ist veränderbar und subjektiv

Kompetenz kann nicht als konstante Kategorie angesehen werden, da sie eine personengebundene Kategorie darstellt. Unter diesem Gesichtspunkt wird von ihrer Veränderbarkeit und Entwicklungsfähigkeit ausgegangen (vgl. Wollersheim, 1993, S. 97), z. B. beim Erlernen von Lesen oder Schreiben. Bei fehlenden Anwendungsmöglichkeiten ist ein Kompetenzverlust vorstellbar, da vorhandene Potentiale weder angewendet noch weiterentwickelt werden können (vgl.

1. Sprachkompetenz und Sprachkompetenzmodelle

Lempert, 1998, S. 140; Pietrzyk, 2002). Eine solche Situation ist beim Erlernen einer Fremdsprache denkbar: Verwendet man die Fremdsprache über eine längere Zeit hin nicht, kann es zu Kompetenzverlusten kommen.

Neben Kompetenzverlusten sind auch Kompetenzverschiebungen zwischen Kompetenzdimensionen denkbar. Kanning (2003) vertritt die Meinung, dass eine zeitbezogene Erfassung von Kompetenz nicht hinreichend sei, da so bestenfalls eine Zustandsbeschreibung von Kompetenz erreicht werden könne, nicht aber eine Abbildung von Kompetenzentwicklungsabläufen. In diesem Zusammenhang weist die Autorin darauf hin, dass bei zufälligen Handlungen grundsätzlich nicht von einem Vorhandensein von Kompetenz ausgegangen werden kann. Dazu ist vielmehr eine mehrfache Erfassung von Kompetenz notwendig, die sich zu verschiedenen Zeitpunkten ähnelt und damit nicht dem Zufall zuzuschreiben ist.

Durch diese Auffassung werden viele Verfahren zur Erfassung von Sprachkompetenz in Frage gestellt, da sie sich häufig auf eine Momentaufnahme beziehen und diese als Kompetenz interpretieren. Die Momentaufnahme stellt im diesem Zusammenhang allerdings nur einen Ausgangspunkt dar, der nicht als eine konstante Kategorie angesehen werden soll, sondern als ein Punkt in einem Entwicklungsablauf. Es gibt Testkonstrukteure, die an dem Entwicklungsprozess interessiert sind und entsprechende Testverfahren einsetzen, z. B. Testungen am Anfang und Ende des Schuljahres, um den Entwicklungsprozess zu beobachten.

Die vorliegende Studie hat nicht das Ziel, Entwicklungsprozesse zu beschreiben, sondern fokussiert sich eher auf den Aufnahmepunkt, um den Sprachstand in eben diesem Moment zu bestimmen und zu analysieren. Fokussiert werden, wie bereits erwähnt, Lesefähigkeiten der Studierenden, die sich mit verschiedenen Lesetests erfassen und messen lassen. Sie werden in dieser Studie als basale Fähigkeiten angesehen, die zur Entwicklung schriftsprachlicher Kompetenz beitragen und so insgesamt die Sprachkompetenz wachsen lassen (dazu Kap. 1.5); ferner können sie wesentlich zum Studienerfolg beitragen, wenn man hohe Sprachkompetenz als eine Voraussetzung für die Studierfähigkeit betrachtet (dazu Kap. 3.4).

4. Kompetenz ist von der Zielsetzung abhängig

Die Erfassung von Kompetenz ist immer im Zusammenhang mit den ihr zugrunde liegenden Zielen zu betrachten. Bei jeder Kompetenzerfassung bleibt demnach zu fragen, wer diese mit welchen Intentionen vornimmt und wie die erhobenen Daten interpretiert werden. Oft wird die Komplexität des Gegenstandes nicht beachtet oder unterschätzt, sodass die Erfassung einseitig und stark vereinfacht diskutiert wird (vgl. Kaufhold, 2006, S. 26f.).

Zusammenfassend lässt sich festhalten, dass die Kompetenz einer Person durch die Aktivierung vorhandenen Wissens in Form von Fähigkeiten und Fertigkeiten in bestimmten Handlungssituationen sichtbar und messbar wird. Vor der Erfassung von Kompetenzen muss der Handlungsbezug definiert und exakt beschrieben werden.

Im Folgenden soll der Begriff *Sprachkompetenz* näher definiert werden, da es in dieser Untersuchung um die Messung der Sprachfähigkeiten von Studierenden geht. Die in Kap. 1.1 dargestellten Kompetenzbegriffe werden um weitere für die Sprachtesttheorie relevante Definitionen erweitert. Kap. 1.3 geht stärker auf die sprachlichen Aspekte der Kompetenzforschung ein.

1.3 Zum Begriff *Sprachkompetenz*

Für den Begriff *Sprachkompetenz* gibt es, ebenso wie für den Begriff Kompetenz, keine einheitliche Definition, da in vielen Disziplinen der Schwerpunkt auf unterschiedlichen Aspekten liegt. Da der Sprachkompetenzbegriff in die Konstruktion verschiedener Sprachkompetenzmodelle eingeflossen ist, ist zunächst ein Überblick über Definitionen von Sprachkompetenz notwendig und eine Vorstellung verschiedener Sprachkompetenzmodelle (Kap. 1.5).

Der Begriff Sprachkompetenz bezieht sich auf ein sehr komplexes Phänomen. Sprachkompetenz liegt ein lebenslanger Entwicklungsprozess zugrunde, bei dem sich der Lerner phonologische, lexikalische und pragmatische Charakteristika der Sprache aneignet (vgl. Klein, 2000, S. 549). Während ein Lerner in der

1. Sprachkompetenz und Sprachkompetenzmodelle

Muttersprache in der Regel alle vier Teilfertigkeiten (Hören, Sprechen, Lesen, Schreiben) an die jeweiligen Kulturstandards angepasst erwirbt, können im Fremdsprachenunterricht einzelne Fertigkeiten auch getrennt ausgebildet werden. Am häufigsten wird Lesefertigkeit isoliert vermittelt, ohne dass die anderen drei Fertigkeiten trainiert werden.

Der Gemeinsame Europäische Referenzrahmen (GER), der die Beherrschung der Sprache (Fremdsprache) auf verschiedenen Stufen beschreibt, definiert die diesbezüglichen Kompetenzen als

„... Summe des deklarativen Wissens, der prozeduralen Fertigkeiten und der persönlichkeitsbezogenen Kompetenzen und allgemeinen kognitiven Fähigkeiten“ (Europarat, 2001, S. 21).

Die Sprachverwendenden werden in diesem Kontext als ‚sozial Handelnde‘ betrachtet. Kognitive Fähigkeiten und persönlichkeitsbezogene Kompetenzen (z. B. Temperament, Motive, Erfahrungen etc.) haben großen Einfluss darauf, wie eine Person in der vorgegebenen Situation (z. B. Kommunikationssituation) handeln wird.

Die vorliegende Studie setzt sich das Ziel, sprachliche Fähigkeiten und Fertigkeiten von Studierenden unter einem linguistischen Aspekt zu untersuchen. Daher soll im Weiteren verstärkt auf den Sprachkompetenzbegriff in den Bereichen Linguistik und Psycholinguistik fokussiert werden.

Linguistik betrachtet Sprache als ein System komplexer Strukturen und beschäftigt sich mit der Beschreibung sprachlicher Phänomene, z. B. Grundelemente der Sprache: Laute, Morpheme, Wörter, Sätze und Texte. Linke, Nussbaumer & Portmann (2001) definieren Sprachkompetenz handlungsbezogen in der Verwendung von Sprache als Kommunikationsmittel, unter Einbeziehung der grundsätzlichen Regeln zum Sprachgebrauch.

Die Psycholinguistik befasst sich mit Prozessen des Spracherwerbs, der Sprachproduktion und des Sprachverstehens auf der Suche nach Erklärungen, wie Sprache im menschlichen Gehirn repräsentiert und verarbeitet wird. Sie betrachtet Sprachkompetenz mit dem Fokus auf kognitiven Vorgängen bei der Rezeption und Produktion sprachlicher Äußerungen (vgl. Hartig, 1999).

1. Sprachkompetenz und Sprachkompetenzmodelle

In der vorliegenden Studie wird Sprachkompetenz als Beherrschung der grundlegenden sprachlichen Elemente (Phonetik, Morphologie, Syntax, Lexik, Semantik) verstanden, die je nach Sprache (Erst-, Zweit- oder Fremdsprache), Alter des Lernenden und Zweck des Erlernens unterschiedlich stark ausgebildet sein können. Im Anschluss an die Definition von Kompetenz von Carle (2003) wird auch hier zwischen individuellen und kollektiven Kompetenzen unterschieden. Da bei Sprachstandserhebungen angestrebt wird, die Fähigkeiten eines Individuums zu identifizieren, um z. B. sprachliche Schwächen zu diagnostizieren, wird hier die Rolle des Individuums oder des Handelnden unterstrichen.

Ein Individuum (der Lerner, der Sprachverwendende) hat verschiedene Ressourcen zur Verfügung, die es im Spracherwerbsprozess erworben hat und die es je nach Handlungssituation einsetzt. Es bestimmt, welche sprachlichen Fertigkeiten und Fähigkeiten in einer bestimmten Handlungssituation aktiviert werden müssen, um eine zu erwartende oder anzustrebende Leistung erfolgreich zu erbringen.

1.4 Entwicklung des Sprachkompetenzbegriffs

Im Folgenden wird auf die Entwicklung des Sprachkompetenzbegriffs eingegangen, der in verschiedene Sprachkompetenzmodelle eingeflossen ist und im Rahmen dieser Arbeit für weitere Fragestellungen von Interesse ist. Die Aspekte, die untersucht werden, sind für das Verständnis der Verschiebung der Begriffsgrenzen wichtig.

Den Grundstein zur Definition von Sprachkompetenz in der Linguistik legte Ferdinand de Saussure zu Beginn des 20. Jahrhunderts mit seinem Werk „Cours de linguistique générale“. Er unterschied zwischen „*langue*“, einem statischen System von Formen und Strukturen, die als Wissen oder Kognition zu verstehen sind, und „*parole*“ als Realisierung des Wissens im Sprechakt.

Chomsky (1957, 2001) griff diese Unterteilung auf, benannte diese Begriffe aber anders: „*langue*“ wurde zu *Performanz*, und „*parole*“ zu *Kompetenz*. Bei ihm war

1. Sprachkompetenz und Sprachkompetenzmodelle

Kompetenz nunmehr kein statisches Wissen mehr, sondern ein dynamisches Konzept, das Regeln und Operationen miteinander verband. Zur Ausführung dieses sprachlichen Wissens wurde Performanz notwendig (vgl. Jude, 2008, S. 13).

Die Frage nach dem Handlungsbezug überträgt sich vom Kompetenzbegriff auf die Sprachkompetenzforschung und wird vor allem in der Linguistik als Problem von Kompetenz und Performanz diskutiert, wobei einer pädagogisch-psychologischen Kompetenzerfassung besondere Aufmerksamkeit geschenkt wird (vgl. Jude, 2008, S. 12). Bis heute wird diskutiert, was erfasst wird: Kompetenz oder Performanz.

Hymes (1971) dehnte den Begriff Kompetenz auf die kommunikative Ebene aus und prägte in diesem Zusammenhang den Begriff der *kommunikativen Kompetenz* (vgl. Dresselhaus 1979, S. 170). Mit ihm sollten soziale Grundfähigkeiten besondere Berücksichtigung finden, die eigens erworben werden und für die Kommunikation erforderlich sind (vgl. Glück 2000, S. 356).

Die Trennung zwischen sprachlichem Wissen und sprachlichem Verhalten als Anwendung dieses Wissens führte zu Debatten in der empirischen Sprachforschung. Hymes beschrieb die Trennung folgendermaßen: „competence is understood to be dependent on two things: (tacit) knowledge and (ability for) use“ (Hymes 1971, S. 16). Diskutiert wurde u. a. auch, ob sich die sprachliche Kompetenz, so wie sie nach Chomsky erfasst wurde, mit Testverfahren erfassen ließe, wenn diese nur die Performanz erfassen könnten, also die Anwendung des Wissens und nicht die Kompetenz selbst (vgl. Jude, 2008, S. 13).

Die Trennung zwischen theoretischem Sprachwissen und praktischem Sprachhandeln manifestierte sich seit den sechziger Jahren in vielen Theorien der Sprachkompetenz (dazu Cummins, 1979; Morrow, 1979; Canale & Swane, 1980; Bachman & Palmer, 1982; Bachman, 1990). Der Fokus wurde dabei auf kommunikative Kompetenz und Performanz gelegt.

Mit der Entwicklung der kommunikativen Didaktik in den siebziger Jahren fand der Sprachkompetenzbegriff Verwendung in der Fremdsprachendidaktik und floss in die Entwicklung von Kompetenzstufen für Fremdsprachen ein, wo er im Gemeinsamen Europäischen Rahmen (GER) sehr umfangreich erfasst wurde.

1. Sprachkompetenz und Sprachkompetenzmodelle

Auslöser der Diskussion um die kommunikative Kompetenz im Fremdsprachenunterricht in der Bundesrepublik Deutschland war vor allem das 1974 erschienene Buch von H-E. Piepho *Kommunikative Kompetenz als übergeordnetes Lernziel im Englischunterricht*.

In der Definition von Sprachkompetenz bestehen unter anderem Unterschiede auch dadurch, dass Sprachkompetenz in Abhängigkeit von der Sprache, die untersucht wird (Erst-, Zweit- oder Fremdsprache), in Teilkomponenten oder Teilkompetenzen zerlegt wird.

Im Folgenden werden verschiedene Modelle vorgestellt, die Sprachkompetenz in der Erst- und Fremdsprache erfassen, für die dann eigene Kompetenzbeschreibungen erstellt werden.

1.5 Modelle zur Erfassung von Sprachkompetenz

Zur Beschreibung der Kompetenz werden verschiedene Sprachmodelle und sprachtheoretische Ansätze (vgl. dazu Kap. 2.2) ausgearbeitet und als Grundlage für die Interpretationen verwendet. Sprachmodelle werden dabei sehr fein differenziert und beschreiben viele Dimensionen und Facetten, die nicht zu verallgemeinern sind. Dabei gelten in den Studien, die Sprachkompetenzen untersuchen: a) unterschiedliche Rahmenbedingungen, b) verschiedene Testziele und c) Verfahren zur Messung der Sprachkompetenz, die oft untereinander nicht vergleichbar sind.

Die Sprachvermittlung in der Erstsprache zielt immer auf Sprache als Ganzes ab, jedoch werden bestimmte Teilbereiche stärker fokussiert.

Modelle für die Erfassung von Fremdsprachenkompetenz gehen davon aus, dass die Kompetenz sich aus verschiedenen Komponenten zusammensetzt (vgl. Lado, 1961; Harris, 1969). Das Modell von Lado (1961) unterscheidet z. B. sechs Elemente: Aussprache, Intonation, Betonung, Grammatik, Wortschatz und Übersetzung, die sich in den vier Grunddimensionen Lesen, Schreiben, Sprechen und Hören niederschlagen. Harris (1969) ordnet diesen vier Dimensionen die

1. Sprachkompetenz und Sprachkompetenzmodelle

Komponenten der Phonologie, der Struktur, des Wortschatzes und der Sprachflüssigkeit zu. Er unterscheidet ferner die Prozesse des Decodierens und Encodierens sowie gesprochene und geschriebene Sprache.

Auf der Unterscheidung von diesen Grunddimensionen basieren viele Theorien und theoretische Ansätze zur Sprachkompetenzstruktur. Diese Dimensionen werden entweder stärker differenziert oder zu übergeordneten Faktoren zusammengefasst. Sie liegen den empirischen Studien sowohl für die Erstsprache als auch für die Fremdsprache zugrunde. Die vier Grunddimensionen bilden eine Grundlage für die übergeordneten Faktoren: Es wird zwischen *Sprachmodalitätsebene* und *Prozessebene* differenziert. Sie werden in auditive (Hören und Sprechen), schriftsprachliche (Lesen und Schreiben), produktive (Sprechen und Schreiben) und schließlich rezeptive (Hören und Lesen) Kompetenzen unterteilt (s. Abb. 1).

Abb. 1: Modell der vier Grunddimensionen von Sprachkompetenz

		Modalität	
		Gesprochen auditiv	Geschrieben visuell
Prozess	Enkodieren produktiv	Sprechen Grammatik Wortschatz Lautsystem Flüssigkeit	Schreiben Grammatik Wortschatz Rechtschreibung
	Dekodieren rezeptiv	Hörverstehen Grammatik Wortschatz Lautsystem	Lesen Grammatik Wortschatz Geschwindigkeit

Quelle: Jude, 2008, S. 34

Empirische Studien, die sprachliche Dimensionen untersuchen, lassen sich grob in zwei Gruppen einordnen:

- 1) Studien, die nur *einen* übergeordneten Faktor empirisch untersuchen
- 2) Studien, deren Ergebnisse *mehrdimensionale* Modelle unterstützen

Welches Modell zur Erfassung von Sprachkompetenz gewählt wird, hängt von den Testinstrumenten ab, die in der jeweiligen Studie zum Einsatz kommen. Ein Beispiel für eine Studie, in der Sprachkompetenz eindimensional erfasst wurde, stützt sich auf die von Oller (1976) formulierte Annahme. Oller vertritt die These,

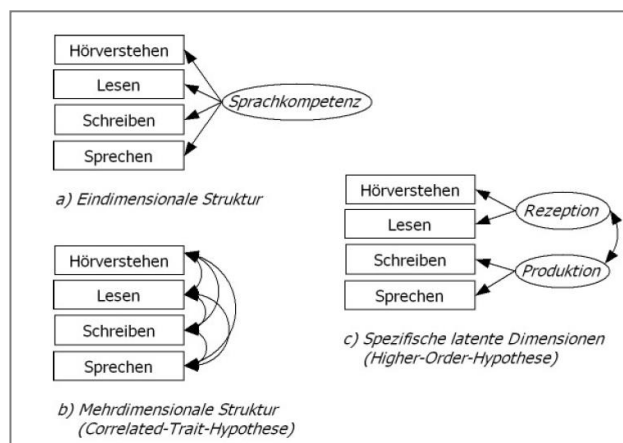
1. Sprachkompetenz und Sprachkompetenzmodelle

dass der sprachliche Informationsverarbeitungsprozess auf einem grundlegenden Faktor, einem Generalfaktor, basiert. Bei der Sprachverwendung wird gleichzeitig auf verschiedene Teilkompetenzen zurückgegriffen. Als Folge dieser Hypothese wurden Sprachinstrumente entwickelt, die Sprache global erfassen sollten und keine Differenzierung sprachlicher Teilbereiche vorsahen (vgl. Raatz, 1985; Feldmann, Grotjahn & Stemmer, 1986). Auch den C-Tests, die dieser Studie als Hauptinstrument zur Erfassung sprachlicher Fähigkeiten und Fertigkeiten von Studierenden zur Verfügung stehen, liegt diese Idee zugrunde.

Die Ergebnisse von Oller, die auf der Grundlage exploratorischer Analyse entstanden und globale Sprachfähigkeit mit einem einzigen Faktor, der Varianz, erklärten, wurden stark kritisiert (vgl. Sang & Vollmer, 1980). So beschäftigte sich die Fremdsprachenforschung mit mehrdimensionalen Modellen, die Sprachkompetenz stärker differenzierten. Es gab zwei Annahmen, die in verschiedenen Studien überprüft wurden: Die *Correlated-Trait-Hypothese* und die *Higher-Order-Hypothese* (s. Abb. 2).

Bei der Correlated-Trait-Hypothese ist die Sprachkompetenz auf mehrere Traits (Fähigkeiten) zurückzuführen, die sich als separate, jedoch korrelierende Faktoren abbilden lassen. Nach der Higher-Order-Hypothese werden auch verschiedene Traits unterschieden; sie werden jedoch zu einem oder mehreren Faktoren zusammengefasst (vgl. Jude, 2008, S. 38).

Abb. 2: Darstellung der eindimensionalen und mehrdimensionalen Modelle zur Sprachkompetenzstruktur



Quelle: Jude, 2008, S. 38

1. Sprachkompetenz und Sprachkompetenzmodelle

Studien, denen die oben beschriebenen Hypothesen zur Sprachkompetenzstruktur zugrunde liegen, unterscheiden sich sowohl in ihren Testverfahren als auch in den untersuchten Alters- und Sprachgruppen und sind deswegen nicht vergleichbar. Die sprachlichen Dimensionen werden nur vereinzelt erhoben und analysiert. In der empirischen Forschung finden sich in der Regel Befunde zum Zusammenhang zwischen zwei oder maximal drei sprachlichen Teilbereichen wieder, da die Erfassung aller Teilkompetenzen gleichzeitig sehr zeit- und kostenaufwendig ist (z. B. TestDaF, TOEFL).

Häufig werden in Studien zur Sprachkompetenz produktive Sprachkompetenzen erhoben, die mit Lese- und Schreibkompetenz verbundene Fertigkeiten abbilden. Das lässt sich dadurch erklären, dass sowohl Lesen als auch Schreiben institutionell vermittelt und oft auch evaluiert werden. Für solche Fertigkeiten wie Hörverstehen und Sprechen soll in der Regel mit der Einschulung ein funktionales Niveau erreicht sein, das durch Schulreifeuntersuchungen, die auch mit Sprachtests verbunden sein können (vgl. z. B. Delfin 4), überprüft wird. Nach der Einschulung werden Hörverstehen und Sprechen in der Muttersprache meist nicht mehr erhoben. Hier wird von der idealistischen Annahme einer funktionalen Kompetenz ausgegangen, die es ermöglicht, dem Unterricht sprachlich zu folgen. Da dies nicht immer der Fall ist, werden in zunehmendem Maße auch in den Schulen sprachliche Fördermaßnahmen angeboten (vgl. Jude, 2008, S. 38).

Die Kontrolle der Entwicklung des Hörverstehens und der Sprechfertigkeit spielt jedoch eine wesentliche Rolle im Fremdsprachenerwerb (vgl. Jude, 2008). Es gibt eine Reihe von Untersuchungen, die unterschiedliche Teilbereiche erfassen und untereinander oder mit verschiedenen Faktoren korrelieren (vgl. Bachman & Palmer, 1981; Bachman, 1990; Eckes, 2000; North, 2000).

Im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* an der Universität Duisburg-Essen, in dem auch die vorliegende Studie entstanden ist, werden die schriftsprachlichen Fähigkeiten von Lehramtsstudierenden gemessen. Überprüft wird, ob sich die Schreibfähigkeit, erfasst in einer separaten Schreibaufgabe, der ein komplexer argumentativer Text zugrunde liegt, durch das Screening-Instrument *C-Test* vorhersagen lässt (vgl. Bremerich-Vos et al.,

2013). Die Fähigkeit des Test-Instruments C-Test, allgemeine Aussagen über die Sprachkompetenz der getesteten Personen zu machen, wurde in vielen Untersuchungen nachgewiesen (vgl. hierzu Kap. 4).

Das Ziel der vorliegenden Studie, in der C-Tests zusammen mit Lesetests verglichen werden, ist es, zu überprüfen, ob sich durch den Einsatz adressatenspezifischer C-Tests und Lesetests ähnliche Teilkompetenzen von Studierenden messen und in gleichem Maße abbilden lassen (vgl. dazu Kap. 6, Kap. 7).

Im Folgenden soll auf zwei große Studien eingegangen werden, die im Rahmen ihres Messkonzepts die Sprachkompetenz von Muttersprachlern erfassen.

1.6 Vorstellung ausgewählter Studien zur Erfassung von Sprachkompetenz

Zwei große Schülerleistungsstudien, auf die ausführlich eingegangen werden soll, sind die PISA- und die DESI-Studie, die zu den sogenannten *Large-Scale-Assessments* zählen. An ihrem Beispiel soll gezeigt werden, wie die Sprachkompetenz von Muttersprachlern erhoben wird und welche Teilkompetenzen dabei im Fokus stehen.

Für die vorliegende Arbeit sind beide Studien von besonderer Relevanz, da Sprachkompetenz in ihnen sowohl global als auch differenziert erfasst wird. Im Weiteren wird verstärkt auf Ziele, Methoden und Kompetenzbereiche eingegangen.

Beide Studien entstanden auf der Grundlage einer sehr umfangreichen Stichprobe⁴ und erheben den Anspruch auf Repräsentativität für ein ganzes Bildungssystem. Im Rahmen solcher Studien erfolgt meist eine Auswahl und Beschränkung auf bestimmte sprachliche Kompetenzbereiche, die erfasst werden sollen. In der Regel spielt bei der Auswahl der Testinhalte im schulischen

⁴ In der DESI-Studie wurden ca. 11.000 Schüler getestet.

1. Sprachkompetenz und Sprachkompetenzmodelle

Bereich die Anbindung an die Curricula oder Bildungsstandards eine zentrale Rolle⁵.

Die genannten Studien befassten sich verstärkt mit der Erfassung unter anderem von sprachlichen Kompetenzen, allerdings mit unterschiedlichen Methoden, denen verschiedene Grundkonzepte zur Verfügung standen.

Die **PISA-Studie** steht für "*Programme for International Student Assessment*" und setzt sich zum Ziel, basale Kompetenzen von fünfzehnjährigen Schülern in den Bereichen Lesekompetenz (*Reading Literacy*), mathematische Grundbildung (*Scientific Literacy*) und fächerübergreifende Kompetenzen (*Cross-Curricular Competencies*) zu erfassen.

Die **DESI-Studie** oder "*Deutsch Englisch Schülerleistungen International*" untersuchte einerseits die sprachlichen Leistungen der neunten Jahrgangsstufe im Englischen und Deutschen, andererseits wurden Curricula, Lehrtexte und Unterrichtsmaterialien und vor allem Unterrichtsgestaltung revidiert.

1.6.1 PISA-Studie

Ziel und Zielgruppe: Bei PISA werden (fach-)sprachliche Fähigkeiten von Jugendlichen (Fünfzehnjährigen) untersucht. Für die Untersuchung dieser Fähigkeiten wurden folgende Dimensionen festgelegt, die nicht fachspezifisch sind und über die Orientierung an den schulischen Curricula hinausgehen: Erfassung von Lesekompetenz sowie von mathematischer Grundbildung und naturwissenschaftlicher Bildung. PISA erhebt den Anspruch, „über die Messung von Schulwissen hinauszugehen und die Fähigkeit zu erfassen, bereichsspezifisches Wissen und bereichsspezifische Fertigkeiten zur Bewältigung von authentischen Problemen einzusetzen.“⁶

Für die vorliegende Studie sind diese Annahmen zur Struktur der Lesekompetenz von Bedeutung; daher wird hier besonders auf diesen Kompetenzbereich näher eingegangen. Die Entwicklung der Lesekompetenz nach PISA bildet einen

⁵ Ob und inwieweit es gelingt, die in den Bildungsstandards formulierten Lernziele zu erreichen, wird regelmäßig in mehrjährigen Abständen vom IQB überprüft und berichtet. Die Ergebnisse werden dabei für die einzelnen Länder in der Bundesrepublik Deutschland dargestellt, <http://www.iqb.hu-berlin.de/bista/control>, zuletzt gesehen am 26.06.2013.

⁶ Grundkonzeption des internationalen PISA-Projektes, <http://www.mpib-berlin.mpg.de/Pisa/intgrundkonzeption.htm>, zuletzt gesehen am 05.03.2013.

1. Sprachkompetenz und Sprachkompetenzmodelle

zentralen Grundstein für die Entwicklung verschiedener fachsprachlicher Kompetenzen und nimmt im Gesamtkonzept der allgemeinen Sprachkompetenz einen bedeutenden Stellenwert ein (vgl. Abb. 3).

Grunddefinition der Lesekompetenz:

Der Lesekompetenz liegt folgende Definition zugrunde:

„die Fähigkeit, geschriebene Texte unterschiedlicher Art in ihren Aussagen, ihren Absichten und ihrer formalen Struktur zu verstehen und sie in einen größeren sinnstiftenden Zusammenhang einzuordnen, sowie in der Lage zu sein, Texte für verschiedene Zwecke sachgerecht zu nutzen“ (PISA 2000, S. 22).

Nach Weinert (1999) kann Lesekompetenz als Handlungskompetenz bezeichnet werden, da sie nicht nur kognitive Leistungskomponenten, sondern auch Haltungen, Einstellungen und notwendige Strategien umfasst. Leser werden hier als Akteure gesehen, die u. a. mithilfe sprachlicher Mittel im sozialen Kontext situations- und kontextadäquat *handeln* können.

Lesekompetenz bedeutet nach der PISA-Konzeption, nicht nur in der Lage zu sein, mithilfe eines Textes Verständnisfragen zu beantworten (z. B. wie im Multiple-Choice-Test), sondern eine sinnvolle Textrepräsentation im Gedächtnis aufzubauen⁷. Um die Lesekompetenz differenziert zu erfassen, wird zwischen textimmanenten und wissensbasierten Verstehensleistungen differenziert. Solange der Text dem Schüler vorliegt, werden *textimmanente* Verstehensleistungen erfasst. Die im Text enthaltenen Informationen bieten eine ausreichende Grundlage für die Beantwortung der Fragen. Wenn kein Zugriff auf den Text möglich ist, muss eine situationsadäquate Interpretation unter Rückgriff auf nicht im Text enthaltenes Vorwissen entwickelt werden, was laut PISA-Definitionen *wissensbasierte* Verstehensleistungen darstellt.

Unterschieden wird in der PISA-Studie zwischen verschiedenen Textsorten, Anwendungssituationen und Leseaufgaben, die unterschiedliche Aspekte des Textverständnisses erfassen. Lesekompetenz wird hier als ein Werkzeug zum

⁷ In der nationalen Ergänzungsstudie zu PISA wurden daher zusätzlich Texte eingesetzt, um verschiedene Verstehensleistungen zu erfassen. Es wurde zwischen der Bearbeitung eines Textes vor der Testphase und während der Lesephase unterschieden (vgl. PISA 2000, S. 24f., auch Untersuchungen von Kintsch, 1994, 1998 und van Dijk & Kintsch zur Theorie des Textverstehens, 1983).

1. Sprachkompetenz und Sprachkompetenzmodelle

Handeln im Sinne der Kompetenzauffassung von Weinert verstanden. In Tabelle 1 sind Anforderungen an die Lesekompetenz formuliert, die mit verschiedenen Leseaufgaben gemessen werden.

Tabelle 1: Darstellung der PISA-Dimension (Lesen)

Bereich	Lesekompetenz
	Geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potential weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen.
Komponenten/ Dimensionen des Bereichs	<p>Verschiedene Arten von Texten lesen: kontinuierliche Texte, klassifiziert nach Typen (z. B. Beschreibung, Erzählung) und Dokumenten, klassifiziert nach Struktur.</p> <p>Verschiedene Arten von Leseaufgaben ausführen, etwa bestimmte Informationen heraussuchen, eine Interpretation entwickeln oder über den Inhalt oder die Form eines Textes reflektieren.</p> <p>Texte lesen, die für verschiedene Situationen geschrieben wurden, z. B. für persönliche Interessen oder um Arbeitsanforderungen zu genügen.</p>

Quelle: Deutsches PISA-Konsortium, 2000, Tabellenauszug: *Zusammenfassende Darstellung der PISA-Dimensionen*, S. 23.

Messinstrumente: Zur Überprüfung des Text- und Leseverständnisses wurden halboffene und geschlossene Aufgaben verwendet (vgl. Kap. 5). Das Produkt, das am Ende des Leseprozesses steht, wird als *Textverständnis* definiert. *Leseverständnis* ist in diesem Zusammenhang das Resultat einer aktiven Auseinandersetzung mit dem Geschriebenen. Die erzielte Leistung hängt von verschiedenen Merkmalen des Lesers und des gelesenen Textes ab. Neben den basalen Lesefähigkeiten gehören auf Seiten des Lesers vor allem kognitive Grundfähigkeiten, Sprach-, Welt- und inhaltliches Wissen sowie motivationale Faktoren wie Wertorientierung und Interessen dazu, bestimmte Fähigkeiten und Fertigkeiten zu aktivieren (vgl. dazu das Kap. 2).

Die Aufgaben, die den Lesetests zugrunde liegen, bestehen aus Diagrammen und unterscheiden sich von den klassischen Lesetexten, die eine kontinuierliche, lineare Struktur aufweisen. Anschließend an Grafiken oder Diagramme werden Aufgaben gestellt in Form halboffener Fragen mit einer Multiple-Choice-Auswahl

1. Sprachkompetenz und Sprachkompetenzmodelle

oder offene Fragen, bei denen die Schüler ihre Antworten selbst produzieren müssen (vgl. Beispielaufgaben aus dem Lesekompetenztest, PISA 2000)⁸.

Das Konzept von Lesekompetenz in der PISA-Studie ist verbunden mit dem Nachweis einer textbasierten Handlungsfähigkeit der Schüler. In einem modulierten Kontext sind sie dazu aufgefordert, mit den ihnen zur Verfügung stehenden sprachlichen Mitteln zu handeln, d. h. Texte zu lesen, notwendige Informationen herauszusuchen und über das Gelesene zu reflektieren. Unter Handlung wird hier eine aktive Auseinandersetzung mit sprachlichen Kategorien und Bereichen verstanden. Kommt der Schüler mithilfe verschiedener Leistungskomponenten und notwendiger Strategien zu einem guten Testresultat, bedeutet dies, dass er erfolgreich gehandelt hat.

Lesetests, wie sie in der PISA-Studie zum Einsatz kommen, liegen für die Gruppe der Studierenden nicht vor. Die Entwicklung entsprechender Tests ist äußerst aufwändig. Deshalb wird das Leseverständnis im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* an der Universität Duisburg-Essen mit anderen Lesetests gemessen, denen adressatenspezifische Textsorten zugrunde liegen und die zeitökonomisch durchzuführen und auszuwerten sind. Im Studium bildet Lesekompetenz die Grundlage zur Entwicklung weiterer Teilkompetenzen, z. B. Schreibkompetenz. Für universitäre Anforderungen eignen sich argumentative Texte, die eine hohe Informationsdichte aufweisen und wissenschaftlichen Mustern entsprechen. Im Laufe der Studiums muss die Fähigkeit weiterentwickelt werden, Argumentationsstrukturen in Lesetexten zu erkennen und Informationen herauszufiltern und in einen eigenen Text einzubauen, sodass er wissenschaftlichen Strukturen entspricht.

In der vorliegenden Studie bietet sich zur Überprüfung der Lesekompetenz der C-Test an (Gründe zur Auswahl des Testformats s. Kap. 4).

Zur Lösung der C-Tests ist das Leseverständnis erforderlich, was nach der PISA-Definition das Resultat einer aktiven Auseinandersetzung mit dem Geschriebenen darstellt. Im Sinne der C-Test-Konstruktion bedeutet dies, dass

⁸ PISA 2000, Beispielaufgaben aus dem Lesekompetenztest, <http://www.oecd.org/berlin/39803735.pdf>, zuletzt gesehen am 26.06.2013.

1. Sprachkompetenz und Sprachkompetenzmodelle

eine komplette Rekonstruktion eines C-Tests nur nach einer aktiven Auseinandersetzung mit verschiedenen in den Text eingebauten Lücken möglich ist. Der C-Test-Text lässt aufgrund mehrerer Lücken kein fließendes Lesen zu, das linear verläuft, sondern erfordert verschiedene Lesestrategien (dazu ausführlich in Kap. 6). Für die Bereiche Wortschatz, Kollokationen, Grammatik und Syntax werden bei der Rekonstruktion von C-Tests Kenntnisse und Fähigkeiten auf entsprechenden sprachlichen Ebenen benötigt. Durch die Rekonstruktion von C-Tests lassen sich demnach Rückschlüsse auf Leseprozesse auf verschiedenen Ebenen ziehen (dazu in Kap. 6), und aufgrund der sprachlichen Beschränkung sind Rückschlüsse auf automatisierte Prozesse und die Verarbeitungsgeschwindigkeit möglich, die beim kontextuellen Schließen der Lücken eine wichtige Rolle spielen (vgl. dazu ausführlich Wockenfuß, 2009).

Bezogen auf die vorliegende Studie wird die Kompetenz, die mit Hilfe von C-Tests erfasst wird, folgendermaßen definiert: Je kompetenter die Probanden im Deutschen sind, desto besser können sie die C-Tests lösen, indem sie „im Zuge einer Rekonstruktion und antizipatorischen Verarbeitung von Sprache Gebrauch von der natürlichen Redundanz eines Textes machen“ (Lehmann, Gänßfuß & Peek, 1999, S. 23). Prozesse, die durch den Einsatz von C-Tests erfasst werden, werden wie folgt beschrieben:

„A C-Test measures the ability to apply and integrate contextual, semantic, syntactic, morphological lexical and orthographic information and knowledge pertaining to a particular written language.“ (Hasting, 2002, S. 66)

Die morphologische und orthografische Ebene spielt beim Einsatz der C-Tests für Studierende keine Rolle (dazu ausführlich in Kap. 4). Daher werden C-Tests als Testinstrument u. a. zur Erfassung der Lesefertigkeit in Verbindung mit verschiedenen Teilkompetenzen (Wortschatz sowie Lesestrategien und Möglichkeit des Rückschließens auf automatisierte Prozesse der Sprachverwendung) eingesetzt.

C-Tests lassen allgemeine, globale Aussagen bezüglich der sprachlichen Fähigkeiten der Probanden zu, weshalb sie in der Praxis oft als Screening-Instrument genutzt werden (dazu Kap. 4). Auch in der vorliegenden Studie

1. Sprachkompetenz und Sprachkompetenzmodelle

wurden C-Tests als Instrument zur Überprüfung der allgemeinen sprachlichen Kompetenz und speziell auch der Lesekompetenz verwendet.

1.6.2 DESI-Studie

Ziel und Zielgruppe: In der DESI-Studie wurden die sprachlichen Leistungen von Schülern der neunten Jahrgangsstufe in den Fächern Deutsch und Englisch untersucht. Das Kompetenzkonzept stellt einen direkten curricularen Bezug im Bereich der Messung von Schülerkompetenzen dar und unterscheidet sich damit vom Lesekompetenz-Konzept der PISA-Studie. Sprachliche Kompetenzen wurden bei DESI zu Beginn und zum Schuljahresende getestet, um einen Zuwachs an Kompetenzen innerhalb der neunten Jahrgangsstufe ermitteln zu können. Sprachliche Fähigkeiten der Schüler in den Fächern Deutsch und Englisch wurden mithilfe unterschiedlicher Testaufgaben überprüft und summierten sich aus verschiedenen Teilbereichen bzw. Teilkompetenzen (s. Tabelle 2).

Kompetenzbeschreibung: Für die Fächer Deutsch und Englisch wurden Kompetenzen formuliert und beschrieben, die in Zusammenarbeit mit Experten aus verschiedenen Bereichen (Fachdidaktik, Testentwicklung, Schulforschung) in Kompetenzmodelle umgesetzt wurden.

Für das Fach *Deutsch* wurden folgende Kompetenzen beschrieben, die sich auf detaillierte Lehrplananalysen stützen: Lesekompetenz, Bewusstheit für grammatische und stilistische Phänomene, Rechtschreibung, Wortschatz, Verständnis für Argumentationsmuster und zwei Aspekte der Schreibkompetenz (semantischer/pragmatischer Bereich, d. h. adressatengerechte Wortwahl und Textkonstruktion sowie sprachsystematischer Bereich, d. h. Satzkonstruktion, Orthografie, Zeichensetzung).

Die Tests im Fach *Englisch* orientieren sich an den Anforderungen des Gemeinsamen Europäischen Referenzrahmens (GER). Für das Fach Englisch wurden folgende Kompetenzen formuliert: mündliche Sprechfähigkeit, Hörverstehen, Leseverstehen, kreatives Schreiben, zwei Aspekte der Sprachbewusstheit (grammatischer Bereich und sozio-pragmatischer Bereich, d. h. Verständnis für adressatengerechte Sprache), interkulturelle Kompetenz

1. Sprachkompetenz und Sprachkompetenzmodelle

und Fähigkeit zur Rekonstruktion von Lückentexten auf der Grundlage des C-Tests.

In Bezug auf die Lesekompetenz konzentriert die DESI-Studie sich auf das Verständnis von literarischen Texten und Sachtexten, da diese im Unterricht häufig verwendet werden. Hier besteht, wie bereits erwähnt, der Hauptunterschied zur PISA-Studie (vgl. Klieme, 2007).

Im Folgenden wird auf die Testkonzeption der DESI-Studie eingegangen, die von den KMK-Fachexperten mit den KMK-Standards für die 9. Klasse im Fach Deutsch und Englisch verglichen wurde. In den Tabellen 4 und 5 werden diese Bereiche ausführlich erläutert.

In Tabelle 2 werden in der ersten Spalte die zu messenden Teilkompetenzen beschrieben, in der Spalte *Testkonzeption* das der DESI-Studie zugrunde liegende Konzept erfasst, und in der dritten Spalte werden Einschätzungen bezüglich der curricularen Gültigkeit für das Fach Deutsch angegeben. Unabhängige Experten aus den Fachkommissionen der Kultusministerkonferenz (KMK) wurden um eine fachliche Einschätzung und Bewertung von DESI-Aufgaben zur Erarbeitung von Bildungsstandards gebeten. In der Spalte Aufgabenwichtigkeit wurde von ihnen die Relevanz bestimmter Kompetenzdimensionen sowohl für den Unterricht als auch in Bezug auf das Erreichen des Mittleren Schulabschlusses (10. Klasse) geschätzt. Auffällig bei der Einschätzung der Aufgabenwichtigkeit ist, dass der Rechtschreibung weniger Wichtigkeit (67 %) zugesprochen wird als dem Wortschatz (96 %).

Tabelle 2: Bestandteile des DESI-Tests (Fach Deutsch)

Bestandteile des DESI-Tests	Testkonzeption	Aufgaben wichtig/sehr wichtig (in %)
Sprachbewusstheit (Deutsch)	Grammatische und semantische Fehler in Sätzen erkennen und Sätze in korrigierter Form aufschreiben; Konjunktivformen erkennen und korrekter Umgang mit direkter/indirekter Rede; Stilformen erkennen und schriftsprachlich umschreiben	85 %
Argumentation	Einschätzung einer Situation und erste kommunikative Reaktion; Einschätzung einer argumentativen Aussage; Produktion von Argumenten, basierend auf (einem) Wissen; argumentative Machart der Begründung	85 %

1. Sprachkompetenz und Sprachkompetenzmodelle

Lesekompetenz	Sinntragende Wörter im Text zuverlässig finden; Inferenzen bilden und Fokus auf schwierige Stellen richten; Öffnen von Wissen und Verknüpfen von Textstellen; innere Repräsentation wesentlicher Textaspekte haben	91 %
Rechtschreibung	Beherrschung der Schriftsprache; Lautstruktur der Wörter erkennen, entsprechend gliedern und in nachvollziehbarer Form verschriftlichen; Beherrschung der orthografischen Regeln; Beherrschung von Schreibungen, die weder durch Regeln noch Analogien erschlossen werden können	67 %
Textproduktion	Offizielle und/oder persönliche Briefe zu vorgegebenen Themen schreiben	91 %
Wortschatz	Umgang mit Worten aus dem Grundwortschatz, mit weniger gebräuchlichen Abstrakta und Konkreta und mit Fach- und Fremdwörtern	96 %

Quelle: DESI-Konsortium, 2008, Tabellenauszug: *Ergebnisse des Expertenratings für die Deutsch-Tests*, S. 28.

Nach der oben aufgeführten Tabelle der Kompetenzen sollen die Schüler erst den Grundwortschatz beherrschen und ausweiten und sich erst dann mit den Regeln zur Schriftsprache beschäftigen. Wenn orthografische Regeln nicht beherrscht und angewendet werden, können die Schüler den erlernten Grundwortschatz, über den sie verfügen (Fach- und Fremdwörter, weniger gebräuchliche Konkreta und Abstrakta), nicht korrekt verschriftlichen, was wiederum negative Auswirkungen auf die Bewertung von sprachlichen Leistungen haben kann.

Im Folgenden wird auf die Darstellung der Kompetenzen und Aufgaben für das Fach Englisch eingegangen, weil im Aufgabenpool der DESI-Studie auch die C-Tests Verwendung gefunden haben.

Tabelle 3: Bestandteile des DESI-Tests (Fach Englisch)

Bestandteile des DESI-Tests	Testkonzeption	Aufgaben wichtig/sehr wichtig (in %)
Leseverstehen	Informationen (konkrete bis abstrakte) in alltäglichen Kontexten erkennen; diese verknüpfen und interpretieren können, um Hauptaussagen zu verstehen; abstrakte Informationen verknüpfen oder inhaltlich komplexe Einzelinformationen interpretieren können	89 %
Hörverstehen	Einzelinformationen aus Kontexten alltäglicher Kommunikation (einfacher bis komplexerer) hörend	

1. Sprachkompetenz und Sprachkompetenzmodelle

	verstehen; konkrete Informationen beim Hören verknüpfen und ein Verständnis der Hauptpassagen entwickeln; implizite Informationen aus Gehörtem erschließen oder explizite Informationen interpretieren können	84 %
Textproduktion (schriftlich)	Offizielle und/oder persönliche Briefe zu vorgegebenen Themen schreiben können	75 %
Textrekonstruktion (C-Test)	Allgemeine Kompetenz (genereller Sprachstand) in der Fremdsprache Englisch; Lücken in einem Text mit den korrekten Worten in der richtigen grammatikalischen Form füllen können	0 %
Interkulturelle Kompetenz	In vorgegebenen Interaktionssituationen ("Critical Incidents") relevante kulturelle Unterschiede erkennen und erfahren sowie in kulturangemessenere Form denken und handeln können	91 %
Sprachbewusstheit (Grammatik)	Fehler bzgl. (einfacher bis komplexer) morphologisch-syntaktischer Regelungen, Tense und Aspects erkennen und verbessern können	100 %
Sprachbewusstheit (Pragmatik)	Soziokulturelle und textpragmatische Regelungen des Englischen anwenden und Verstöße oder Unregelmäßigkeiten erkennen können; sprachliches Handeln durchschauen	100 %

Quelle: DESI-Konsortium, 2008, Tabellenauszug: *Ergebnisse des Expertenratings für die Englisch-Tests*, S. 31.

Im Fach Englisch betonen die Experten die Wichtigkeit der Sprachbewusstheit (pragmatischer und grammatischer Bereich), während der Testkonstruktion in Form von *C-Tests* eine geringere Wichtigkeit zugesprochen wird. Diese Einschätzung wird im Folgenden kommentiert.

In Kombination mit sechs anderen Testaufgaben im Fach Englisch wurden 12 verschiedene C-Tests als Globalindikator der Sprachkompetenz verwendet (vgl. Klieme, 2008). Das Testinstrument *C-Test* wurde von den auswärtigen Experten am häufigsten dem KMK-Kompetenzbereich *Verfügung über sprachliche Mittel/Wortschatz* zugeordnet. Die Einschätzung der Aufgaben durch die Experten ergab, dass der C-Test folgendem KMK-Bildungsstandard zugeordnet wurde: „Schülerinnen und Schüler sind in der Lage [,] zusätzliche lexikalische Einheiten hörend oder lesend zu verstehen oder selbstständig aus Texten zu erschließen.“ (DESI, 2008, S. 31). Problematisch erscheint jedoch die Zuordnung von C-Tests dem Kompetenzbereich *Textrekonstruktion*. Dies deutet darauf hin, dass Aufgaben- bzw. Aufgabenformate und Kompetenzbereiche in der DESI-Studie vermischt werden. Die folgende Auflistung macht deutlich, dass zwischen folgenden Kompetenzen unterschieden wird: *Lese- und Hörverstehen*,

1. Sprachkompetenz und Sprachkompetenzmodelle

Textproduktion, Textrekonstruktion (C-Test), interkulturelle Kompetenz, Sprachbewusstheit (Grammatik) und Sprachbewusstheit (Pragmatik). Fasst man die Kompetenzen in dieser Form zusammen, wie das beim *Leseverstehen* oder *Hörverstehen* der Fall ist, müsste man die C-Tests dem Kompetenzbereich *globale Sprachkompetenz* statt *Textrekonstruktion* zuordnen, da Rekonstruktion den Bearbeitungsprozess beschreibt und nicht die Kompetenz selbst.

Bei der Kompetenz *Textproduktion (schriftlich)*, die ebenso wie der C-Test keine Kompetenzbeschreibung, sondern eine Operationalisierung bzw. den Bearbeitungsprozess darstellt, wird nicht zwingend nur die Produktion eines Textes überprüft, sondern es können u. a. auch Leseverständnis, Wortschatz oder Argumentationsstruktur fokussiert und miterfasst werden.

Der C-Test hat sich in der Forschung als ein Testinstrument etabliert (vgl. dazu Kap. 4), das zur *Erfassung der allgemeinen Sprachkompetenz* oder – wie im Rahmen der vorliegenden Studie – zur *Überprüfung von Lesefähigkeiten (Lesekompetenz)* eingesetzt wird. Analysiert man die Ergebnisse des Expertenratings genauer, könnte man den C-Test ebenso gut dem Bereich *Leseverstehen, Sprachbewusstheit (Grammatik) oder Pragmatik* zuordnen. Alle diese Bereiche würde der C-Test im Rahmen seiner globalen Erfassung abbilden. Allerdings liefert der C-Test, je nachdem, in welcher Sprache und mit welcher Zielgruppe er eingesetzt wird, zu den genannten Kompetenzbereichen nur eingeschränkt diagnostisch verwertbare Werte. Daher wird er in der Regel als Globalindikator zur Einschätzung der Sprachkompetenz verwendet (vgl. dazu ausführlich Kap. 4).

Die DESI-Experten erklären die geringe Wichtigkeit des C-Tests in der Testkonzeption der Studie, die im Vergleich zu den anderen Aufgaben den C-Tests zugesprochen wurde, dadurch, dass das Aufgabenformat von den KMK-Experten als unterrichtsfremd und daher kritisch angesehen wird, was auch als höfliche Umschreibung dafür gelesen werden kann, dass die KMK-Experten den C-Test und seine Qualitäten nicht kennen.

Trotz dieser Geringschätzung des C-Tests durch die KMK-Experten wird der C-Test in der DESI-Studie verwendet. Durch die Rekonstruktion von C-Tests können laut der Beschreibung von DESI Rückschlüsse auf die allgemeine Kompetenz und

1. Sprachkompetenz und Sprachkompetenzmodelle

den allgemeinen Sprachstand der Schüler in der Fremdsprache gezogen werden. Bei der Rekonstruktion von C-Tests werden unter anderem grammatisches und semantisches Wissen aktiviert, da Lücken im Text mit korrekten Worten in der richtigen grammatikalischen Form gefüllt werden müssen. Das zugrunde gelegte Kompetenzkonzept von DESI bzgl. der C-Tests geht davon aus, dass

„... umso mehr Lücken gefüllt werden, je kompetenter die Schülerinnen und Schüler in der Fremdsprache sind: Durch Wortelidierungen reduzierte sprachliche Redundanzen können durch antizipatorische Sprachverarbeitung aufgelöst und die fehlenden Worthälften mittels sprachlichen und außersprachlichen Wissens rekonstruiert werden. Je automatisierter die Sprachverarbeitung abläuft, desto schneller dürften die Lücken zu füllen sein“ (DESI, 2002).

Dass der C-Test von den KMK-Experten als irrelevant bewertet wurde, könnte damit verbunden sein, dass der C-Test häufig kritisiert wurde (vgl. zu Kritik am C-Test das Kap. 4.7 und die Publikation von Baur & Mashkovskaya, 2013).

Der durch die KMK-Experten geäußerten Skepsis dem C-Test gegenüber kann an dieser Stelle nicht zugestimmt werden, da im Rahmen des schulischen Fremdsprachenunterrichts in der 9. Klasse der C-Test eine wichtige Rolle spielt. Zum Beispiel können durch den Einsatz von C-Tests Teilkompetenzen mitbewertet (z. B. Wortschatzkenntnisse) und so der Leistungszuwachs (z. B. zum Anfang und zum Ende des Schuljahres) festgestellt werden (vgl. Kap. 4). Wie in der DESI-Studie berichtet wird, besitzen C-Tests eine hohe curriculare Validität, weil die C-Test-Texte auf linguistischen und textuellen Anforderungen der Curriculumanalysen basieren; sie sind außerdem auf den Horizont der Zielgruppe hin abgestimmt und erfassen eine große Bandbreite von Themen (vgl. Harsch & Schröder, 2007, S. 214).

Die Konstruktion von C-Tests wurde mit den darauf anschließenden Analysen der Textgrundlagen beschrieben – zum einen in Bezug auf die Komplexität der Texte auf semantischer Ebene, zum anderen in Bezug auf den Lösungsprozess und die anzuwendenden Strategien (z. B. Einbeziehung von Weltwissen und Interpolationstechniken). Wenn man diese Ebenen bewertet, erfasst der C-Test viel mehr, als im Bereich der Textrekonstruktion sowohl in der Testkonzeption als auch in den KMK-Bildungsstandards beschrieben wird (vgl. DESI, 2008, S. 31).

1. Sprachkompetenz und Sprachkompetenzmodelle

Befunde aus der DESI-Studie in Bezug auf die C-Tests:

In der DESI-Studie wurden anhand der vorliegenden Daten Leistungsniveaus formuliert, die Übereinstimmungen mit den Niveaus des Gemeinsamen Europäischen Referenzrahmens (GER) aufweisen. Die DESI-Niveaus⁹ zeigen Berührungspunkte mit den GER-Niveaus A1 bis C1, doch bei DESI erfolgte keine empirische Überprüfung der Niveaustufen (vgl. Harsch & Schröder, 2008, S. 149f.).

Die Befunde zum globalen Sprachstand der Schülerinnen und Schüler der neunten Jahrgangsstufe durch C-Tests zeigten, dass es hinsichtlich der fremdsprachlichen Leistungen zwischen Bildungsgängen bzw. Schultypen bedeutende Unterschiede gibt. So sind mehr als ein Viertel der Lernenden in Gymnasium und Realschule auf einem unteren Niveau oder darunter angesiedelt. Diese Lernenden können laut den von DESI vorformulierten Leistungsniveaus

„... vertraute, hochfrequente Phänomene verarbeiten und verfügen über elementarste Kenntnisse; von einer Automatisierung der Sprachverarbeitung kann jedoch nicht die Rede sein“ (Harsch & Schröder, S. 155).

Die Ergebnisse an Hauptschulen zeigten sogar, dass die Mehrheit der Schüler das unterste Kompetenzniveau nicht erreichte. Harsch & Schröder gehen davon aus, dass

„... die Lernenden in diesen Bildungsgängen über ein so wenig ausgeprägtes Sprachvermögen verfügen, dass sie auch einfachste fremdsprachliche Anforderungen mehrheitlich nicht erfüllen“ (S. 156).

Im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* an der Universität Duisburg-Essen wird der C-Test aufgrund seiner Gütekriterien verwendet (vgl. Kap. 4.6, Kap. 4.8). Mit dem C-Test wird die allgemeine Sprachkompetenz unter besonderer Berücksichtigung der schriftsprachlichen Kompetenzen der Lehramtsstudierenden gemessen. Ziel der vorliegenden Studie ist es, zu überprüfen, ob durch adressatenspezifische C-Tests auf der einen Seite und adressatenspezifischen Lesetests auf der anderen Seite ähnliche

⁹ In der DESI-Studie wurde der Wortlaut der Deskriptoren von GER mit den Beschreibungen der DESI-Niveaus verglichen.

1. Sprachkompetenz und Sprachkompetenzmodelle

Teilkompetenzen von Studierenden gemessen werden (vgl. dazu Kap. 6 und Kap. 7).

2. Lesekompetenz

In diesem Kapitel soll die Frage geklärt werden, was unter dem Begriff *Lesekompetenz* zu verstehen ist. Durch die Vorstellung verschiedener Lesemodelle soll verdeutlicht werden, welche Kompetenzkomponenten existieren und welche von ihnen durch entsprechende Tests messbar sind.

2.1 Was ist Lesekompetenz?

Der Begriff Lesekompetenz wird in der wissenschaftlichen Literatur und der öffentlichen Diskussion unterschiedlich verwendet. Um den Verwendungskontext im Rahmen dieser Studie einzugrenzen, wird hier nur auf Definitionen eingegangen, die für die vorliegende Arbeit relevant sind.

In der PISA-Studie wird der Lesekompetenzbegriff in einem handlungsorientierten Kontext positioniert und fokussiert die pragmatischen Elemente von Lesekompetenz. Ein kompetenter Leser ist danach einer, der mit Informationen, die er aus Texten zu entnehmen in der Lage ist, am gesellschaftlichen und politischen Leben teilnehmen kann (vgl. PISA, 2000, S. 78).

Der kompetente Leser ist in der Lage,

„... geschriebene Texte zu verstehen, zu nutzen und über sie zu reflektieren, um eigene Ziele zu erreichen, das eigene Wissen und Potential weiterzuentwickeln und am gesellschaftlichen Leben teilzunehmen“ (PISA, 2000, S. 23).

Lesekompetenz ist also nicht nur die Fähigkeit, schriftliche Informationen zu entziffern, sondern auch eine „aktive (Re-)Konstruktion der Textbedeutung“, die sowohl mit sprachlichem Wissen als auch mit Vor- und Weltwissen verbunden ist (PISA-Konsortium, 2000, S. 71).

Diese Definition liegt sowohl der PISA- und IGLU-Studie als auch dem OECD-*Programm for the International Assessment of Adult Competencies* (PIAAC) zugrunde, das mit dem Ziel initiiert wurde, das Kompetenzniveau von Erwachsenen im internationalen Vergleich zu untersuchen¹⁰.

¹⁰ PIAAC wird in den 26 beteiligten Ländern als Querschnittuntersuchung in mehreren Wellen durchgeführt. Die erste Welle startete 2008, die Ergebnisse werden 2013 publiziert.

2. Lesekompetenz und Lesekompetenzmodelle

Die Leseforschung ist sich heute weitgehend darin einig, dass Lesen ein komplexer Vorgang ist, der mehrere Teilprozesse umfasst und über die Fertigkeit, Texte zu decodieren, hinausgeht (vgl. Kintsch, 1992; Christmann & Groeben, 1999). Nicht jeder dieser Teilprozesse lässt sich durch Lesetests beobachten und messen. Die Teilprozesse verbinden Wort-, Satz- und Textebene, sodass im Text semantische, syntaktische und textübergreifende Relationen hergestellt werden können. Artelt et al. (2007) beschreiben diese Prozesse folgendermaßen:

„Auf der untersten Ebene sind das Erkennen von Buchstaben und Wörtern und die Erfassung von Wortbedeutungen angesiedelt (Wortebene), auf der mittleren Ebene steht die Herstellung semantischer und syntaktischer Relation zwischen Sätzen im Vordergrund (Satzebene). Auf der höheren Ebene geht es dann um die satzübergreifende Integration von Sätzen zu Bedeutungseinheiten und den Aufbau einer kohärenten mentalen Repräsentation der Bedeutung des Textes (Textebene)“ (Artelt et al., 2007, S.11).

Die dargestellten Ebenen, die Teilkomponenten der Lesekompetenz ausbilden, werden von Weinert (2001) durch solche vervollständigt, die sich auf die individuellen Merkmale der Leser (z. B. Fähigkeit, Wissen, Können, Handeln, Erfahrung und Motivation) fokussieren und auf die beim Lesen zugegriffen werden muss. Auf sie gehen in ihrem Mehrebenen-Lesemodell auch Rosebrock & Nix (2008) ein. Die Prozessebene wird danach durch zwei weitere Ebenen vervollständigt: die Subjektebene und die soziale Ebene. Im Ansatz von PISA spielen die von Rosebrock & Nix und Weinert herausgearbeiteten Komponenten keine Rolle.

Will man Lesekompetenz messen und möchte man wissen, wie geschriebene Texte verstanden und genutzt werden und vor allem, wie über sie reflektiert wird, ist die Erfassung einzelner Prozessebenen erforderlich, die nicht immer direkt beobachtbar sind. Die Leseleistung hängt von den Leseanforderungen ab und von individuellen Merkmalen des Lesers selbst. Diagnostische Informationen bleiben manchmal vage (z. B. in den PISA-Studien oder IGLU-Studie) – besonders dann, wenn es darum geht zu wissen, an welchen Stellen leseschwache Leser Schwierigkeiten haben. Die Kompetenzkomponenten und Prozessebenen, die

2. Lesekompetenz und Lesekompetenzmodelle

beim Lesen eine wichtige Rolle spielen, können sehr mannigfaltig sein, wenn wir einen Blick auf komplexe Modellierungen des Leseprozesses werfen, wie sie zum Beispiel von Artelt et al. (2007) entwickelt werden (dazu ausführlich in Kap. 2.1.1, 2.1.2 und 2.1.3).

Da im Rahmen dieser Studie davon ausgegangen wird, dass zur Lösung der eingesetzten Testformate mehrere Ebenen des Leseprozesses durchlaufen werden, die durch unterschiedliche Lesetests gemessen werden können, sollen im Folgenden drei Lesemodelle vorgestellt werden, die Leseprozessebenen beschreiben, die sich auch in den Bearbeitungsprozessen beim Lösen von C-Tests wiederfinden.

Aus dem oben Gesagten kann festgehalten werden, dass Lesekompetenz kein monolithisches Konstrukt ist, sondern sich vielmehr aus miteinander interagierenden Teilfertigkeiten zusammensetzt. Leseverstehen resultiert aus der Wechselwirkung von Verarbeitungsprozessen auf allen Ebenen. Diese Annahme geht auf Goodmans Leseprozessmodell (1967) zurück, das an dieser Stelle genannt werden muss, da sein Modell als Grundlage für die Entwicklung späterer elaborierter Lesemodelle angesehen werden kann, die im Weiteren dargestellt werden. Für Goodman besteht der Leseprozess darin, Hypothesen zu konstruieren und sie danach mithilfe eines Textes zu überprüfen. Für die Beschreibung von Prozessen, die sich bei der kognitiven Verarbeitung von *cues* vollziehen, führte Goodman die Begriffe *sampling*, *predicting*, *testing* und *confirming* ein (vgl. Goodman, 1973, S. 136). Die beschriebenen Prozesse laufen beim geübten Leser in der muttersprachlichen Rezeption automatisiert ab. Sie finden sich sowohl bei fortgeschrittenen und/oder lesestarken Lesern als auch bei Anfängern und/oder leseschwachen Lesern – mit dem Unterschied, dass sie bei fortgeschrittenen Lesern automatisiert und unbewusst ablaufen. Bei Anfängern oder/und leseschwächeren Lernern erfordern sie eine aktive Auseinandersetzung mit verschiedenen sprachlichen Bereichen, deren Entwicklung darin besteht, dass immer mehr Prozesse automatisiert werden.

2. Lesekompetenz und Lesekompetenzmodelle

Die moderne Leseforschung wird von zwei konkurrierende Modellen bestimmt: einem modularen und einem interaktionistischen Ansatz (zur Debatte über verschiedene Modelle vgl. Markmann & Dietrich, 2000).

Modulare Theorien postulieren, dass beim Leseprozess involvierte Teilprozesse autonom ablaufen und voneinander unabhängig sind; höhere Prozesse werden erst dann eingesetzt, wenn die Verarbeitung auf den niedrigen Stufen abgeschlossen ist (vgl. Fodor, 1983; Garfield, 1987). Interaktionistische Modelle hingegen gehen davon aus, dass die Ebenen parallel ablaufen und es bei den Verarbeitungsprozessen zu Überlappungen kommen kann (vgl. das Aktionsmodell von McClelland & Rumelhart, 1981; van Dijk & Kintsch, 1983). Höhere Verarbeitungsprozesse können durchaus vor der Verarbeitung auf den unteren Ebenen durchlaufen werden (vgl. Richter & Christmann, 2002, S. 28).

In der empirischen Forschung haben sich die interaktionistische Modelle als gültig ergeben. Vor allem das Modell von van Dijk und Kintsch (1983) hat sich etabliert und die Text- und Leseforschung stark beeinflusst. In dem von van Dijk & Kintsch beschriebenen Strategiemodell (1983) wird die zuvor entwickelte Propositionsanalyse (Kintsch & van Dijk 1978) ausgeweitet, ergänzt und differenziert.

Nach van Dijk und Kintsch lässt das Lesen sich in Teilprozesse auf verschiedenen Ebene aufgliedern. Die Autoren gehen davon aus, dass sich diese Prozesse weitgehend parallel vollziehen und es zwischen hierarchieniedrigeren und hierarchiehöheren Rückmeldungen gibt. Nach van Dijk & Kintsch (1983) ist die Verständlichkeit, die Lesbarkeit von Texten, nicht nur von den Texten selbst abhängig, sondern ergibt sich auch aus der Interaktion der Textverwender mit den Texten (S. 362). Für das Verstehen eines Textes benötigen die Leser ein Situationsmodell, das ihnen die Vorstellung einer Situation ermöglicht, in der das Textgeschehen ablaufen kann, in der die im Text vorkommenden Personen, Objekte, Orte oder Fakten vorstellbar sind (S. 337). Die Entwicklung des Situationsmodells ist nur auf Grundlage eines durch Erfahrung erworbenen Wissens möglich. Die Lesenden müssen den Text zu ihren Wissensstrukturen in Beziehung setzen, um das für das Textverstehen notwendige Situationsmodell zu entwickeln (S. 337 ff.). Die beschriebenen Prozesse fließen in das Lesemodell von

Richter & Christmann (2002) ein und werden ausführlich in Kap. 2.2.3 beschrieben.

2.2 Lesemodelle

Im Folgenden werden unterschiedliche Lesemodelle vorgestellt, welche die Lesekompetenz aus verschiedenen Perspektiven beschreiben. Die dargestellten Lesemodelle sollen dabei helfen, geeignete Lesetests auszuwählen, die im akademischen Bereich eingesetzt werden können.

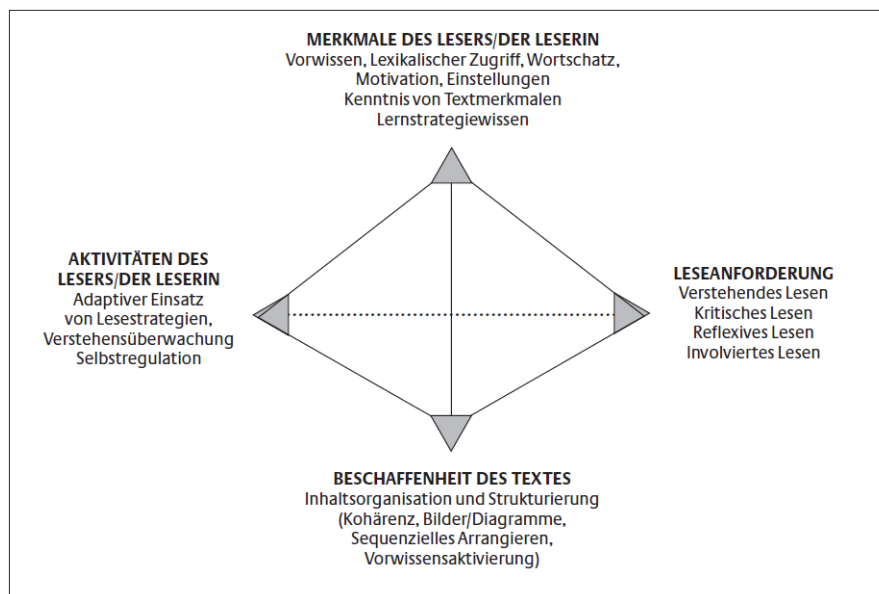
2.2.1 Das Lesemodell von Artelt et al. (2007)

Das Lesemodell von Artelt et al. (2007) verortet verschiedene Komponenten (Determinanten) der Lesekompetenz in einem Koordinatensystem (s. Abb. 6). Die Determinanten beschreiben nicht nur Eigenschaften der Texte, mit denen die Lesenden konfrontiert werden, sondern auch Eigenschaften der Lesenden selbst, da diese individuelle Faktoren den Leseprozess mitbeeinflussen.

Lesefertigkeiten sind von Person zu Person unterschiedlich ausgeprägt und verändern sich im Laufe eines Entwicklungsprozesses, der vom Vorhandensein bestimmter angeborener Persönlichkeitsmerkmale (z. B. Begabung, Intelligenz), kognitiver Merkmale (z. B. Gedächtnis, umfangreiches Wissen, automatisierte Fertigkeiten) und der Bewältigung bestimmter text- und lesebezogener Anforderungen abhängt (Artelt et al., 2007, S. 12).

In Abb. 3 sind die Merkmale dargestellt, die beim Lesen miteinander interagieren und die Auswirkungen auf die Qualität des Textverstehens haben können.

Abb. 3: Determinanten der Lesekompetenz nach Artelt et al. (2007)



Quelle: Artelt et al., 2007, S. 16

1. Merkmale des Lesers/der Leserin

Diese Determinante stellt die individuelle Ebene dar und ist bei einzelnen Lesern unterschiedlich stark ausgebildet. Dazu gehören: Vorwissen, Wortschatz, Motivation, Kenntnis von Textmerkmalen und Lernstrategiewissen von Lesenden.

Ein Besitz von inhaltlich relevantem Vorwissen kann den negativen Effekt schwacher Lesefertigkeiten auf das allgemeine Leseverständnis sogar teilweise kompensieren (vgl. Artelt et al., 2007, S. 13). Mithilfe von vorwissensgestützten Inferenzen ist es möglich, eine zusammenhängende Repräsentation des Textes zu bilden – zum Beispiel ein Verständnis von kausalen Zusammenhängen im Text (vgl. Artelt et al., 2007, S. 14; auch Schnotz, 1994; Singer, 1994).

Die Autoren betonen, dass ein umfangreicher Wortschatz zu einem guten allgemeinen Leseverständnis beiträgt. Entscheidend für den Erfolg von Wortschatztrainings sei die Vermittlung eines breiten, leicht zugänglichen und kontextübergreifenden Wissens über Wortbedeutungen. Solches Wissen erfordere Lehrmethoden, die eine tiefe und vorwissensgestützte Verarbeitung beinhalten, z. B. Wortbedeutungen diskutieren, eigenständige Beziehungen

2. Lesekompetenz und Lesekompetenzmodelle

zwischen neuen und bekannten Wörtern herstellen oder wichtigen Wörtern mehrmals und in unterschiedlichen Kontexten begegnen (vgl. Artelt et al., 2007, S. 15; auch McKeown, Beck, Omanson & Perfetti, 1987).

Ein weiteres Merkmal, so Artelt et al., ist das Wissen über typische Merkmale einzelner Textsorten, das Einfluss auf die Art der Textverarbeitung und das erzielte Textverständnis hat. Kenntnis über Struktur und Aufbau des Textes kann die Rezeption des Textinhaltes vereinfachen. Wissen über die Textgliederung und die Textsorte erhöht das Verständnis, da das vorhandene Wissen bereits aktiviert ist und bestimmte Kohärenzlücken zu schließen hilft. Dadurch wird nicht zuletzt auch die Lesegeschwindigkeit erhöht (vgl. Artelt et al., 2007, S. 15).

Die Motivation, die auch stark von Weinert (2001) bei der Kompetenzbeschreibung betont wird, bezeichnet im Lesekontext das Ausmaß des Wunsches oder der Absicht, in einer bestimmten Situation einen spezifischen Text zu lesen (vgl. Schiefele, 1996). Lesemotivationen können unter anderem Lesegewohnheiten bilden, die wiederum auf andere Ebenen (zum Beispiel der Prozessebene) Auswirkungen haben können – zum Beispiel in Form höherer Leseleistung oder -geschwindigkeit.

In der Forschung wird zwischen *extrinsischer* und *intrinsischer* Motivation unterschieden (vgl. Deci & Ryan, 1985): Entweder bewegen äußere (extrinsische) oder innere (intrinsische) Reize eine Person zum Lesen. Artelt et al. (2007) sind der Meinung, dass intrinsische Motivation den Leseprozess positiv beeinflussen kann und sich dadurch auch auf ein besseres Leseverständnis auswirkt. Die Häufigkeit des Lesens beeinflusst das Wissen in Form von Lesegeschwindigkeit, Leseflüssigkeit, positivem lesebezogenen Selbstkonzept und der damit verbundenen Auswahl anspruchsvoller Texte. Dadurch entsteht eine Automatisierung unterschiedlicher Teilprozesse des Lesens und des Lesehabitus' (vgl. Artelt et al., 2007, S. 19; auch Guthrie et al., 1999).

Wie wichtig das Lernstrategiewissen ist, zeigen die Ergebnisse der PISA-Studie. Den Ergebnissen zufolge korreliert die Lesekompetenz mit dem Lernstrategiewissen¹¹. Das Lernstrategiewissen beinhaltet Wissen darüber, unter

¹¹ Es existiert auch ein korrelativer Zusammenhang zwischen der Leseleistung (und zwar dem Lernstrategiewissen) und den kognitiven Leistungen der Schüler. Der Wert beträgt $r = .55$.

welchen Bedingungen bestimmte Strategien (beim Lesen sind dies Texterschließungsstrategien) erfolgreich eingesetzt werden können (vgl. Artelt, Schiefele & Schneider, 2001).

2. Leseanforderung

Die Leseanforderung beinhaltet verschiedene Modi des Lesens, zum Beispiel verstehendes, kritisches, reflexives oder involviertes Lesen. Sie ergibt sich aus dem Leseausgangstext und der Leseabsicht. Der Ausgangstext mit seiner Spezifik (z. B. Zeitungsartikel, Forschungsbericht oder Instruktion) und die Leseabsicht (z. B. das Heraussuchen von Argumenten für oder gegen eine Hypothese) bestimmen die Auswahl des Lesemodus' (z. B. informierendes Lesen). In den meisten Fällen geht es dabei um die Wechselbeziehung zwischen Leseziel und Leseausgangstext.

Bei der Textverarbeitung gilt das *verstehende* Lesen als die zentrale Qualifizierung des Leseprozesses. Beim verstehenden Lesen werden alle kognitiven und metakognitiven Strategien eingesetzt (Artelt, 2000; Wild, 2000).

Die Informationsverarbeitung ist mit Bewertungsprozessen verbunden, die durch *kritisches Lesen* erreicht werden. Die Binnenstrukturierung des kritischen Lesens greift auf die Konzeption des kritischen Denkens zurück, die auf dem „metatheoretischen wissenschaftlichen Theoretisieren“ aufbaut (Artelt et al. 2007, S. 21). Analyse und Bewertung erfolgen dabei zum Beispiel über Begriffsbedeutungen, logische Strukturen oder inhaltliche Dimensionen.

Das *reflexive* Lesen geht über die im engeren Sinne kognitiven Aspekte der Lesekompetenz teilweise hinaus. Diese Leseart hat mit komplexeren und weiterreichenden Verarbeitungsprozessen zu tun als das verstehende und kritische Lesen, das mit Aspekten der kognitiven und personalen Entwicklung in Verbindung steht. Reflexives Lesen impliziert die Verbindung von Reflexion (z. B. Aufwerfen der Probleme, Hinterfragen von Inhalten und Argumentation) und der kritischen Reflexion (bewusster Denkprozess als Grundlage für die Perspektivenänderung beim Lesen) (vgl. Artelt et al., 2007, S. 22).

2. Lesekompetenz und Lesekompetenzmodelle

Die Qualität des *involvierenden* Lesens bringt emotionale und motivationale Aspekte mit sich, wie zum Beispiel Texte bedürfnisbezogen auswählen, positive Gefühlserlebnisse mit der Lektüre verbinden oder die Lesesituation steuern (vgl. Hurrelmann, 2002, S. 285). Involvierendes Lesen stellt eine Leseform dar, die

„... emotional-motivationale Aspekte akzentuiert, und zwar in Richtung auf Lusterfahrungen[,] von der sinnlichen Anschaulichkeit über Unterhaltungs- und Spannungserleben bis zum empathischen Mitleiden und -freuden“ (Eggert, 2002 zit. in: Artelt, 2007, S. 22).

Die motivierende Funktion dieser Lesevariante ist nicht zu unterschätzen, da involvierendes Lesen durch den damit verbundenen Lustfaktor einen starken Leseanreiz schafft, was für die Leseintensität und die Entwicklung von Lesekompetenz eine nicht zu vernachlässigende Rolle spielt (vgl. Groeben & Hurrelmann, 2002). Dieser Lesemodus ist zunächst mit einer intrinsischen Motivation und einem inhaltsorientierten Interessenkonzept verbunden. Die erworbene Lesekompetenz überträgt sich jedoch später auf andere Textsorten.

3. Beschaffenheit des Textes

Die Determinante *Beschaffenheit des Textes* beinhaltet: Inhaltsorganisation, Strukturierung, Kohärenz und Vorwissensaktivierung des Lesenden in Bezug auf den Lesetext. Lesen stellt in diesem Kontext einen Vorgang dar, bei dem im Optimalfall satzübergreifend eine kohärente mentale Repräsentation des Textes aufgebaut und mit Vorwissen verknüpft wird (vgl. Artelt et al., 2007, S. 23). Die genannten Aspekte überschneiden sich teilweise mit der Determinante Leseanforderung – und zwar mit dem Lesemodus *verstehendes Lesen*, bei dem auch der Einsatz verschiedener Strategien notwendig wäre.

Die kohärente Inhaltsorganisation ermöglicht dem Leser eine reibungslose Textrezeption, indem sie dem Leser Hinweise gibt, welche Sätze und Textteile in welcher Weise aufeinander zu beziehen und in einen sinnvollen Zusammenhang zu bringen sind. Van Dijk & Kintsch (1983) unterscheiden dabei *lokale* und *globale Kohärenzebenen*. Die lokale Kohärenzebene bezieht sich auf die Verknüpfung aufeinander folgender Sätze, globale Kohärenz auf die Integration übergeordneter Textthemen und Textteilen. Fehlen die Verknüpfungshinweise,

so entstehen Kohärenzlücken. Die Argumentation eines Textes ist schneller nachvollziehbar, wenn er auf der Ebene des Textes sprachliche Hilfen anbietet, z. B. wenn X, dann Y (vgl. Christmann & Groeben, 1999; Artelt et al., 2007, S. 24). Die neue Textinformation wird beim Lesen mit dem Vorwissen der Lesenden verknüpft. Dies kann etwa dadurch geschehen, dass der Text mit Erklärungen, Spezifizierungen, Beispielen etc. angereichert ist. Christmann & Groeben sind der Meinung, dass die oben genannten zusätzlichen Informationen lernwirksam sind, weil durch sie ein dichtes Netzwerk von Verbindungen zwischen neuen Informationen und Vorwissen geschaffen wird und ein multipler Zugriff auf das gespeicherte Wissen erfolgt (vgl. Christmann & Groeben, 1996; Artelt et al., 2007, S. 28).

4. Aktivität des Lesers/der Leserin und Lesestrategien

Unter der Determinante *Aktivität des Lesers* wird Selbstregulation als ein adaptiver Ansatz von Lesestrategien und Verstehensüberwachung erfasst.

Effektive Selbstregulation beim Lesen besteht darin, anforderungs- und situationsangemessene Lesestrategien auszuwählen, zu kombinieren und anzuwenden. Die besondere Rolle der Schule in diesem Prozess beschreibt Weinert (1996), indem er das selbstgesteuerte und intentionale Lernen als Voraussetzung, Mittel und Ziel bezeichnet. In der aktuellen Forschung geht man davon aus, dass Lesestrategien konkrete Techniken darstellen, die das Verstehen und Behalten von Textinhalten erleichtern und die zielführend eingesetzt werden können (vgl. Artelt, 2000). Diese Techniken werden entweder automatisiert oder bewusst verwendet. Möchte der Leser inhaltlich wichtige Informationen behalten, kann er sie zum Beispiel im Text farblich markieren oder ihn mit ein paar Stichworten kommentieren. Der weniger kompetente Leser muss diese Technik erst erwerben, bevor er sie automatisiert verwenden kann.

In Abhängigkeit von den jeweiligen Anforderungen beim Textverstehen sind unterschiedliche Lesestrategien relevant. In den KMK-Empfehlungen zum Schulfach Deutsch werden diese Strategien benannt als „mit Texten umgehen“, aber auch „Strategien zum Leseverstehen kennen und anwenden“ können (KMK,

2. Lesekompetenz und Lesekompetenzmodelle

2003, S. 13, 26). Die Lesenden sollen danach über grundlegende Verfahren für das Verstehen von Texten verfügen, selbstständig Informationen aus Texten entnehmen, sie miteinander verknüpfen und mit ihrem Vorwissen verbinden können. Diese Strategien werden der Determinante *Vorwissen* zugeordnet.

Bezüglich der Klassifikation kognitiver Lesestrategien werden in der Literatur üblicherweise folgende Strategien unterschieden: *Wiederholung*, *Organisation* und *Elaboration*. Wiederholungsstrategien unterstützen das Behalten von Textinhalten, zum Beispiel in Form von mehrmaligem Lesen, Auswendiglernen einer bestimmten Textstelle etc. Elaborationsstrategien helfen, Verbindungen zwischen Textinhalten und Vorwissen des Lesenden herzustellen (z. B. konkrete Beispiele, praktische Anwendungen, neue Begriffe auf bereits bekannte beziehen, Zusammenhänge herstellen). Organisationsstrategien helfen, wichtige Informationen zu identifizieren, den Textinhalt zu strukturieren und Verbindungen zwischen verschiedenen Textteilen herzustellen (z. B. Hauptgedanken unterstreichen, Inhalte bildlich veranschaulichen, Gliederungen schreiben) (Artelt et al., 2007, S. 30).

Metakognitive Strategien beziehen sich auf die Kontrolle kognitiver Vorgänge und umfassen drei Strategien: *Planung*, *Überwachung* und *Regulation*. Diese Prozesse sind beim Einsatz verschiedener Lesestrategien von Bedeutung.

Die Planung erfolgt in Form von Formulierung von Zielen und Entwerfen von Fragestellungen zum Lesetest. Die Planungsphase führt dazu, dass für die Erleichterung des Leseverstehens bestimmte Vorwissensbestände aktiviert werden: Zum Beispiel informiert man sich vor dem Lesen eines Artikels über das Thema. Die Überwachungsaktivitäten kontrollieren den Lesevorgang (z. B. Lenkung der Aufmerksamkeit, selbstformulierte Fragen zum Inhalt). Ergibt sich im Leseprozess als Ergebnis der laufenden Überwachung ein bestimmtes Problem, kann zum Beispiel die Lesegeschwindigkeit verringert oder die Textpassage, bei der es zu Verstehensschwierigkeiten kam, wiederholt gelesen werden (Artelt et al., 2007, S. 30).

Gute Informationsverarbeitung beruht vor allem auf der Nutzung von Lern- und Lesestrategien und unterscheidet gute und schlechte Strategeanwender

voneinander (vgl. Borkowski, Schneider & Pressley, 1989, S. 31, zitiert nach Artelt 2007).

„Der gute Strategeanwender hat ein großes Repertoire an Fähigkeiten zum strategischen, aufgaben- und zielbezogenen Lernen, das auch die erfolgreiche Meisterung neuer und komplexer Aufgaben erlaubt. Ihm steht nicht nur ein umfangreiches Wissen über Strategien zur Verfügung, das auch das konditionale Wissen einschließt, welche Strategie je nach Situation und Lernziel wann und wo sinnvoll einzusetzen ist [...], sondern er ist auch in der Lage, die Strategien effektiv einzusetzen. Gleichzeitig verfügt der gute Strategeanwender über die notwendige Motivation, die Strategie auch tatsächlich anzuwenden, da er von deren Nützlichkeit überzeugt ist.“ (Artelt et. al., 2007, S. 31)

Zusammenfassend wäre hervorzuheben, dass sich einzelne Determinanten gegenseitig beeinflussen. Durch die Veränderung eines Merkmals kann das andere Merkmal modifiziert werden. Wenn zum Beispiel ein Fachtext zum Lesen vorgelegt wird (Determinante *Beschaffenheit* des Textes) wird ein anderer Lesemodus angewendet (*verstehendes/kritisches Lesen*) als bei einem literarischen Text (*involvierendes Lesen*) (Determinante *Leseanforderung*).

Die beschriebenen Determinanten zeigen, welche Faktoren die Leseleistung beeinflussen können. Verfügt der Lesende über keine für den Lesetext oder die Lese-Testaufgabe relevanten Strategien oder keine ausreichende Motivation, weil etwa der Text nicht seinem Niveau entspricht und ihn über- oder unterfordert, spiegelt sich das im Textverstehen wider.

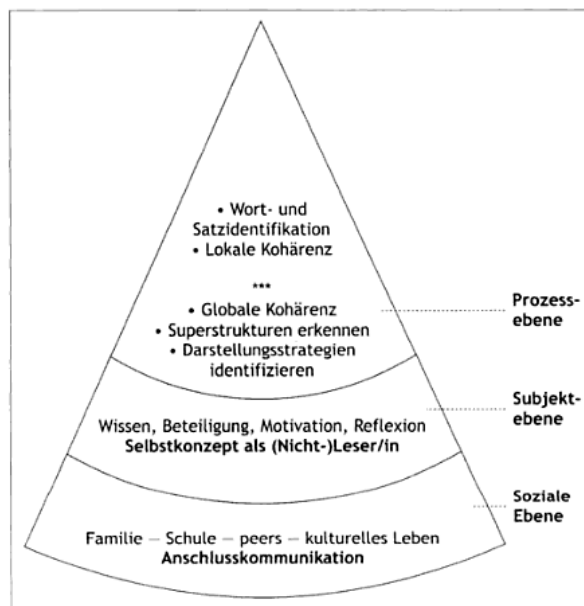
2.2.2 Lesemodell von Rosebrock & Nix (2008)

Rosebrock & Nix (2008) haben das kognitionstheoretische Mehrebenen-Lesemodell entwickelt. Sie betonen die Wichtigkeit des Erwerbs von Lesekompetenz und der subjektiven oder sozialen Funktionen des Lesens in diesem Verlauf. Nach Rosebrock & Nix ist im Lesen ein kognitiv-konstruktiver Vorgang repräsentiert. Die Leistung des kognitionstheoretischen Lesemodells besteht darin, die aktive Tätigkeit des Lesenden in verschiedene Anforderungsdimensionen aufzuschlüsseln, die während des Lesens parallel kognitiv vollzogen werden. In dem dargestellten Modell (Abb. 4) werden drei Ebenen festgelegt: Prozessebene, Subjektebene und soziale Ebene.

2. Lesekompetenz und Lesekompetenzmodelle

Die Prozessebene ist im Rahmen dieser Untersuchung von besonderem Interesse; auf diesen Aspekt soll daher hier näher eingegangen werden. Zur Prozessebene zählen kognitive Fähigkeiten, die hierarchische Prozesse darstellen (vgl. auch van Dijk & Kintsch, 1983; Richter & Christmann, 2002). Kognitive Dimensionen des Lesens sind nicht direkt beobachtbar und können durch Testformate nur teilweise abgebildet werden.

Abb. 4: Mehrebenenmodell von Rosebrock & Nix (2008)



Quelle: Rosebrock & Nix (2008) S. 16.

a) Wort- und Satzidentifikation

Auf dieser Ebene werden Buchstaben identifiziert, zu Wörtern zusammengefasst, und anschließend wird über die Satzstruktur die Bedeutung erschlossen. Die semantische Verfügbarkeit von Wörtern und die kognitive Verarbeitungs- und Lesegeschwindigkeit sind auf der Ebene der Wort- und Satzidentifikation als Prozesse sehr wichtig. Bei Leseanfängern müssen diese Prozesse erst entwickelt werden. Rosebrock behauptet, dass die Leseleistung auf niedrigem Niveau durch einen differenzierten Wortschatz und entsprechendes Kontextwissen entlastet werden kann, indem während des Lesens und Buchstabierens die ‚Top-down‘-Leistung eingesetzt wird. Das bedeutet, dass die Verarbeitung von Wörtern und Sätzen nicht durch einzelne Lexeme und Satzzeichen stattfindet, sondern eher

2. Lesekompetenz und Lesekompetenzmodelle

durch Aktivierung bereits beherrschter Wörtern und über den Kontext des gelesenen Textes.

Für das Satzverstehen ist die aktive Gruppierung von Wörtern und die Herstellung syntaktischer, semantischer und pragmatischer Zusammenhänge zentral. Erfahrene Leser genießen hier einen Vorteil gegenüber Leseanfängern, da sie im rezeptionsbegleitenden Aufbau einer inneren Repräsentation des Gelesenen, eines sogenannten mentalen Modells, geübt sind (vgl. auch Rosebrock, 2007)¹².

b) lokale Kohärenzbildung durch Verknüpfung von Satzfolgen

Rosebrock & Nix vertreten die Meinung, dass sich die genannten kognitiven Prozesse einschließlich des Satzverstehens beim kompetenten Lesen automatisiert vollziehen, ohne Zugriff auf zusätzliche Ressourcen. Für die lokale Kohärenzbildung in Form von Verknüpfungen von Satzfolgen gelte das ebenfalls. Auch bei diesen über die Satzgrenzen hinausgehenden Verknüpfungen spielen das Sprach- und Weltwissen eine bedeutende Rolle. Die Verknüpfung von Satzfolgen konstruiert der erfahrene und kompetente Leser im Vergleich zum Leseanfänger durch Herstellung von Inferenzen völlig unbewusst. Für Leseanfänger oder leseungewohnte Personen seien die Prozesse auf diesen beiden relativ niedrigen Ebenen dagegen mental so aufwändig, dass ihre kognitiven Ressourcen durch diese Anstrengung bereits aufgezehrt werden. Derartige Automatisierungsleistungen auf Wort- und Satzebene bedürfen danach einer langjährigen Erfahrung.

c) Globale Kohärenzherstellung

Von der höheren Ebene Globaler Kohärenzherstellung kann erst dann gesprochen werden, wenn die lokalen Kohärenzen (Verknüpfung von Satzfolgen)

¹² Rosebrock, C. (2007): Reading Literacy und Lesekompetenz. Die kognitive Dimension des Lesens und die innere Beteiligung des Lesers, Artikel vom 23.11.2007, verfügbar unter: <http://www.lesen-in-deutschland.de/html/content.php?object=journal&lid=778>, zuletzt gesehen am 16.05.2013.

2. Lesekompetenz und Lesekompetenzmodelle

hergestellt werden und diese während des Lesens zu den genannten Leistungen der Wort- und Satzidentifikation komprimiert werden. Auf dieser Ebene bildet der Lesende globale, textgestützte Hypothesen, die beim Lesen bestätigt oder verworfen werden. Während des Lesens werden diese Leistungen nach Rosebrock & Nix im mentalen Modell abgebildet, in das neue Informationen aus dem Text Eingang finden, sodass das Modell entweder korrigiert oder differenziert wird. Die Bearbeitung neuer Textelemente durch die Textvorgabe und die Heranziehung weiteren Text- und Weltwissens verläuft parallel, wenn die hierarchieniedrigeren Prozesse sich so automatisiert vollziehen, dass sie nicht die kognitiven Leistungen des Lesenden für sich allein beanspruchen (vgl. auch Rosebrock, 2007).

d) Superstrukturen erkennen

Diese Ebene stellt im Mehrebenenmodell von Rosebrock & Nix die höchste kognitive Prozessebene dar. Wenn die mentalen Modelle in spezifische Texttypen eingeordnet werden können (z. B. rhetorische, stilistische und argumentative Strategien) und entschlüsselt und in ihrer Bedeutung ausgeschöpft werden, bewegt der Leser sich auf der hierarchiehöchsten Ebene des Leseverstehens. Hier kann über den eigenen Leseprozess reflektiert werden. Diese Ebene ist für komplexere Texte notwendig – zum Beispiel das Lesen von wissenschaftlichen Texten.

e) Erkennen von Darstellungsstrategien im Hinblick auf die Textintention

Diese Ebene stellt das Resultat aller vorher erwähnten Prozessebenen dar. Diese Ebene wird von erfahrenen Lesenden beherrscht, welche die Absicht des Textes erkennen und dem Schreibstil des Autors und seiner Intention folgen können. Diese Leseart ist beim Lesen von Texten generell wichtig (unabhängig von der Textsorte) und hilft dem Lesenden zum Beispiel, das Geschriebene besser zu verstehen, indem er sich in die Rolle des Autors versetzt. Auf dieser Ebene spielt

das persönliche Interesse und die Motivation des Lesenden eine große Rolle. Der Lesende entscheidet, ob er den Text verstehen möchte.

Auf diesen fünf kognitiv immer anspruchsvoller werdenden Ebenen werden nach Rosebrock die Textvorgaben vom Leser aktiv als sinnhafte Abbildungen mental repräsentiert und prozessiert (vgl. Rosebrock, 2007; Rosebrock & Nix, 2008, S. 17ff.).

2.2.3 Prozessebenenmodell von Richter & Christmann (2002)

Dem Leseprozessmodell von Richter & Christmann (2002) liegt das Leseprozessmodell von van Dijk & Kintsch (1983) zugrunde, das dem interaktionistischen Modell zugeordnet werden kann, wobei die Teilprozesse hier in hierarchieniedrige und hierarchiehöhere Prozessebenen aufgegliedert werden. Der komplexe Leseprozess wird in verschiedene Ebenen differenziert, die von basalen analytischen Teilprozessen der Buchstaben und Worterkennung über syntaktische und semantische Analyse von Wortfolgen bis hin zum satzübergreifenden Aufbau einer kohärenten Textstruktur reichen.

Mit *hierarchieniedrigen Prozessen* sind die Abläufe während des Lesevorgangs gemeint, die sich auf relativ niedrigem Niveau vollziehen und bei geübten Lesern automatisiert ablaufen. Hierzu zählen basale Wahrnehmungsprozesse, die das Identifizieren und Verarbeiten von Graphemen betreffen. Erst darauf aufbauend können Prozesse auf Wort- und Satzebene entstehen.

*a) Prozesse auf Wortebene: lexikalische Prozesse, phonologische
Rekodierung und Satzkontext*

Prozesse auf *Wortebene* beinhalten sowohl den lexikalischen Zugriff als auch die Nutzung des Satzkontextes bei der Worterkennung (vgl. Richter & Christmann, 2002, S. 29). Dem lexikalischen Zugriff liegt die kognitionspsychologische Vorstellung zugrunde, dass Wortbedeutungen in Form eines mentalen Lexikons im semantischen Gedächtnis niedergelegt sind (vgl. Richter & Christmann, 2002, S. 36). Die zweite mit der Wortebene zusammenhängende Komponente, nämlich

2. Lesekompetenz und Lesekompetenzmodelle

die *Nutzung des Satzkontextes*, bezeugt einen Zusammenhang der Wortidentifikation mit dem jeweiligen Satzkontext. Leser, die in der Lage sind, die korrekte Semantik mithilfe des Satzkontextes herzustellen, zeigen ein höheres Maß an Lesekompetenz (Richter & Christmann, 2002, S. 40).

b) Prozesse auf Satzebene: lokale Kohärenzbildung, syntaktische und semantische Integration

Auf der *Satzebene* wird, zusätzlich zur Wortebene, der gesamte Satz in den Sinnzusammenhang integriert. Dabei muss der Leser die grammatische Struktur des Satzes erkennen. Unterschiede zwischen kompetenten und weniger kompetenten Lesern treten entweder durch Defizite in der syntaktischen Verarbeitung auf oder durch allgemeine Kapazitätsbeschränkung des Arbeitsgedächtnisses. Die Autoren sind der Meinung, dass gute Leser eher in der Lage sind, mehrere Interpretationen syntaktisch ambiger Sätze so lange simultan zu verarbeiten, bis die Mehrdeutigkeit ausgeräumt ist (vgl. Richter & Christmann, 2002, S. 41).

Semantische Integration spielt eine Rolle auf der Satzebene, wenn komplexe Sätze gelesen werden müssen. Mit dem allgemeinen Leseverständnis und der Lesegeschwindigkeit steht sowohl die syntaktische als auch die semantische Integration von Wörtern in Verbindung. Die Bewältigung von Prozessen auf der Satzebene hängt von der individuell verfügbaren Arbeitsgedächtniskapazität ab (vgl. Richter & Christmann, 2002, S. 42)

c) Prozesse auf Textebene: globale Kohärenz, Vorwissen, Arbeitsgedächtnis

Hierarchiehöhere Prozesse werden in der Regel erst dann berücksichtigt, wenn das Leseverständnis (*globale Kohärenz*) überprüft werden soll. Hier wird der Fokus auf die Operationalisierung gelegt, da die Beschreibung der sich dabei vollziehenden Prozesse auf diese Weise verständlicher ist. Die Überprüfung des Leseverständnisses findet in Form von Fragen statt, die mehrere Prozessebenen verbinden: zum Beispiel Kohärenzbildungsprozesse mit basalen Lesefertigkeiten

2. Lesekompetenz und Lesekompetenzmodelle

auf Worterkennungsebene oder auch Top-Down-Prozesse in Form inhaltlichen Vorwissens. Unter globaler Kohärenzbildung verstehen die Autoren alle hierarchiehöheren Prozesse, die zu einem integrativen Verständnis notwendig sind. Die Lesenden müssen in der Regel vorwissensgestützte Inferenzen vornehmen, um zu einer global-kohärenten Repräsentation des Inhalts zu gelangen, da kein Text alle Informationen explizit macht, die zu seinem Verständnis notwendig sind. Das Verarbeitungsergebnis der globalen Kohärenzbildungsprozesse wird „Situationsmodell“ genannt (vgl. Richter & Christmann, 2002, S. 43).

Inferenzen und andere Prozesse auf Textebene sind angewiesen auf die Verfügbarkeit entsprechenden Vorwissens (inhaltliches Vorwissen, Textschemata, pragmatisches Wissen). In „Situationsmodelle[n] sind Repräsentationen des Textinhalts definiert, in denen Textinformation und Vorwissen integriert sind“ (Richter & Christmann, 2002, S. 44, vgl. auch: van Dijk & Kintsch, 1983). Die Autoren nehmen an, dass die Performanz beim Leseverstehen davon abhängt, in welchem Maß verstehensrelevantes Vorwissen individuell verfügbar ist. Besonders bemerkenswert dabei ist, dass Defizite in hierarchieniedrigen Lesefähigkeitskomponenten durch angemessenes inhaltliches Vorwissen vollständig ausgeglichen werden können (vgl. Richter & Christmann, 2002, S. 44; auch: Adams, Bell & Perfetti, 1995; Moravcsik & Kintsch, 1993).

Neben hierarchieniedrigen Prozessen stellt die *Kapazität des Arbeitsgedächtnisses* eine wichtige Komponente dar. Das Kurzzeitgedächtnis erfüllt nicht nur die Funktion, Informationen passiv-temporär zu speichern, sondern es umfasst auch zentrale Verarbeitungsfunktionen.

„Damit kommt dem Arbeitsgedächtnis eine zentrale Rolle für Prozesse der globalen Kohärenzbildung zu, bei denen Teilinformationen an verschiedenen Stellen im Text und Informationen aus dem Langzeitgedächtnis durch aktive Inferenzprozesse integriert werden und das Verarbeitungsergebnis für weitere Inferenzprozesse verfügbar gehalten wird.“ (Richter & Christmann, 2002, S. 45)

Die dargestellten Prozesse auf der Textebene tragen wesentlich zu interindividuellen Unterschieden des Textverständnisses bei. Vorwissen und Kapazität des Arbeitsgedächtnisses haben sich als äußerst wichtige Faktoren für

die Effizienz hierarchiehöherer Prozesse beim Lesen erwiesen. Besonders bemerkenswert ist die Tatsache, dass Defizite in der Bewältigung hierarchieniedriger Prozesse durch effiziente globale Kohärenzbildungsprozesse zumindest teilweise ausgeglichen werden können. Dieser von Richter & Christmann formulierte Befund soll im Folgenden durch den Einsatz von C-Tests, die auf hierarchieniedrigen und hierarchiehöheren Prozessebene messen, überprüft werden (dazu Kap. 6).

2.2.4 Verortung von Tests in Lesemodellen

Im Rahmen dieser Studie werden verschiedene Lesetests zusammen mit C-Tests verwendet, um zu überprüfen, welche Determinanten der Lesekompetenz nach Artelt et al. (2007) gemeinsam abbilden und auf welchen Prozessebenen nach Groeben & Christmann (2002) und Rosebrock & Nix (2008) sie sich bei der Erfassung der Lesefähigkeiten der Studierenden ggf. unterscheiden. Im Folgenden wird versucht, einzelne Lesetests, die in dieser Studie zusammen mit den C-Tests verwendet werden, einzelnen Determinanten zuzuordnen¹³.

Wie bereits beschrieben, fokussieren verschiedene Lesemodelle unterschiedliche Kompetenzkomponenten. Im Modell von Richter & Christmann (2002) werden vor allem die Prozessebenen betont, die den Leseprozess auf verschiedenen Ebenen beschreiben. Im Modell von Rosebrock & Nix (2008) wird außer der Prozessebene noch der Subjektebene und der sozialen Ebene Aufmerksamkeit geschenkt. Das Modell von Artelt et al. (2007) fügt Kompetenzkomponenten hinzu, die neben subjektiven Eigenschaften bzw. Merkmalen der Leser auch Eigenschaften der Texte abbilden.

Da das Modell von Artelt et al. (2007) eine übergeordnete Funktion der Lesekompetenz betont und einzelne Prozessebenen auf dem Koordinatensystem verortet, soll es in Bezug auf die Verortung von C-Tests und Lesetests an dieser Stelle als Erstes beschrieben werden.

¹³ Zum Design der Testung s. Kap. 7.

2. Lesekompetenz und Lesekompetenzmodelle

Die Determinante *Merkmale des Lesers* (Vorwissen, lexikalischer Zugriff, Wortschatz, Kenntnis von Textmerkmalen, Lernstrategiewissen) ist bei der Bearbeitung von allen Tests wichtig. Eine intensivere Auseinandersetzung mit dem Wortschatz in Form lexikalischer, semantischer und syntaktischer Analysen wird sowohl durch den Einsatz von C-Tests und Stolperwörtertests fokussiert als auch durch den Einsatz von MC-Lesetests, da C-Test und Stolperwörtertest zunächst die Bearbeitung auf der Wort- und Satzebene voraussetzen.

Die Beobachtung der Operationalisierung der Prozesse in Form eines Einsatzes von Lesestrategien, von Verstehensüberwachung, metakognitiven Strategien und Selbstregulation, die der Determinante *Aktivitäten des Lesers* zugeordnet sind, lassen sich mit C-Tests und auch mit MC-Lesetests gut erfassen. Sowohl beim C-Test als auch beim MC-Test ist die Auswahl der Lesestrategie wichtig. Bearbeiten die Probanden Lücken im C-Test, ohne den Satz- bzw. Gesamtkontext zu beachten, gelingt es ihnen oft nicht, den gesamten Text zu rekonstruieren. Durch das Suchen nach geeigneten Wörtern im C-Test wird das Verstehen der Sätze überwacht. Gelingt dies nicht, wird nach weiteren Wörtern gesucht. Beim Bearbeiten eines MC-Tests ist auch die Auswahl der Lesestrategie sehr wichtig. Ob die Probanden den Text obeflächlich und schnell oder langsam und genau lesen, hängt (auch) von der Aufgabenstellung ab. Wird eine ungeeignete Strategie gewählt, kann dies Auswirkungen auf das Gesamtergebnis haben. Durch das Lösen der Fragen zum Text wird deutlich, ob die verwendete Lesestrategie und der Lesemodus erfolgreich waren.

Die Aktivierung des Wissens über den Lesetext, Textsorte, Inhaltsorganisation und Strukturierung lässt sich durch den Einsatz von MC-Lesetests abbilden, da diese Texte länger sind und sowohl ein Lesen in Abschnitten und Sequenzen fordern als auch die Arbeit mit der Struktur des Textes, z. B. in Form von selbstformulierten Stichwörtern zu einzelnen Absätzen. Diese Prozesse sind nach Artelt et al. (2007) der Determinante *Beschaffenheit des Textes* zugeordnet.

Die Determinante *Leseanforderung* mit verschiedenen Lesemodi kann durch den Einsatz längerer Texten mit unterschiedlichen Strukturen, unterschiedlicher Thematik und unterschiedlichem Leseziel eingesetzt werden. Dafür eignen sich

2. Lesekompetenz und Lesekompetenzmodelle

MC-Lesetests sehr gut, da sie zum Text spezifische Fragen stellen und den Leseprozess in Bezug auf kritisches oder reflexives Lesen schärfen.

Integriert man die verwendeten C-Tests und Lesetests in das Mehrebenen-Modell von Rosebrock & Nix (2008), werden stark Leseprozesse fokussiert, die die Erfassung der Lesefähigkeiten auf Wort-, Satz- und Textebene ermöglichen. Ähnliche Zuordnungen zu hierarchieniedrigeren und hierarchiehöheren Prozessebenen bietet auch das Modell von Richter & Christmann (2002). Durch die Verwendung dieser Modelle lassen sich C-Tests und Stolperwörtertests gut beschreiben. Daher werden diese Modelle im Weiteren als Grundlage für die Analyse und Kategorisierung der C-Test-Lücken verwendet (vgl. Kap. 6).

3. Sprachtests

Sprachtests sind Testverfahren, die den Sprachstand in der Mutter-, Fremd- oder Zweitsprache überprüfen. Grotjahn (2010b) fasst zusammen:

„... bei Sprachtests im Sinne der pädagogisch-psychologischen Diagnostik handelt es sich um theoretisch und empirisch fundierte Verfahren zur kontrollierten Auslösung von diagnostisch relevantem Verhalten durch standardisierte Reize – mit dem Ziel eines Rückschlusses auf sprachliche Kompetenzen wie Hörverstehens- oder Schreibkompetenz oder auch Wissensstände wie explizite Grammatikkenntnisse.“ (Grotjahn, 2010b, S. 211)

Sprachtests können nach verschiedenen sprachtheoretischen Ansätzen entwickelt und eingesetzt werden (vgl. Klein-Braley, 1981; Raatz, Voss & Klein-Braley, 1991). Das Ziel der Testung und der ausgewählte Ansatz bestimmen das Testformat und die Interpretation der Testergebnisse. Einige wichtige Ansätze werden im Folgenden vorgestellt.

3.1 Ansätze zur Konstruktion von Sprachtests

In der Geschichte der Sprachtests gibt es drei wesentliche Ansätze (vgl. Spolsky, 1985; Wockenfuß, 2009, S. 87), welche die Sprachdiagnostik und Testkultur geprägt haben und bis heute beeinflussen:

- 1) der ‚traditionelle‘ Ansatz
- 2) der psychometrisch-strukturalistische Ansatz
- 3) der psycholinguistisch-soziolinguistische Ansatz

Der ‚traditionelle‘ Ansatz

Nach dem traditionellen Ansatz werden sprachliche Fähigkeiten mit klassischen Methoden (z. B. Aufsatz, Grammatiktests, mündliche Prüfungen, Übersetzungen) überprüft. Häufig ist eine solche Beurteilung recht subjektiv. Die Entwicklung, Auswertung und Interpretation derartiger Sprachtests liegt in den Händen eines

Prüfers bzw. Beurteilers und wird kaum hinterfragt (dazu Süßmilch, 1984a; Raatz, Voss & Klein-Braley, 1991). Sehr häufig, etwa zur Leistungsbeurteilung in Schulen, überprüfen Lehrer mit selbstgestellten Aufgaben und selbst konstruierten Lösungsschlüsseln die Teilkompetenzen ihrer Schüler.

Der traditionelle Ansatz wurde von der Sprachwissenschaft stark kritisiert (vgl. Lado, 1961; Carroll, 1961; Spolsky, 1995). Den Testverfahren wurde vorgeworfen, den Hauptgütekriterien Objektivität, Reliabilität (Verlässlichkeit und Messgenauigkeit) und Validität (Gültigkeit) nicht zu genügen (Ingenkamp, 1995; Süßmilch, 1984). Trotz vieler offensichtlicher Mängel werden traditionelle Testverfahren zur Messung der Sprachfähigkeit auch heutzutage noch eingesetzt und nicht selten als ‚high-stakes‘ verwendet, das heißt in Situationen, die für die Ausbildungs- oder Berufsbiografie entscheidend sind. Damit folgt man vermutlich Forderungen der Praxis (vgl. Wockenfuß, 2009, S. 88).

Der psychometrisch-strukturalistischer Ansatz

Mit dem psychometrisch-strukturalistischen Ansatz versuchten die Testkonstrukteure, die Messgenauigkeit und den Bezug auf linguistische Theorien zu verbessern. Linguistisch wurde die Sprache zu dieser Zeit als ein mehrdimensionales Merkmal aufgefasst, da diese Phase sehr von der Sprachtheorie des amerikanischen Strukturalismus geprägt war. Man ging von strukturell unterschiedlichen Sprachebenen aus, die sich in einzelne abzutestende homogene Teilbereiche einteilen lassen (z. B. Phonologie, Orthografie, Morphologie, Syntax oder Lexikon). Über die verschiedenen Ebenen Hören, Sprechen, Lesen und Schreiben konnten diese Sprachebenen in einzelnen Subtests eines Tests untersucht werden (vgl. Carroll, 1961; Lado, 1961; Harris, 1969; Wockenfuß, 2009, S. 89). Die Leistungen innerhalb der oben genannten Bereiche sollten durch den Einsatz neuer Aufgabenformate noch objektiver messbar werden. So entstanden zum Beispiel Papier-Bleistift-Tests mit Multiple-Choice-Aufgaben, welche die Objektivität der Durchführung und Auswertung erhöhten. Ergänzt wurden diese Formate mit Hörverstehensaufgaben vom Tonband (vgl. Raatz, Voss & Klein-Braley, 1991). Die Einschätzung der

Sprachbeherrschung eines Probanden summierte sich somit aus einzelnen Subtests.

In Bezug auf den psychometrisch-strukturalistischen Ansatz wurde verstärkt die Validität in Frage gestellt (vgl. Oller, 1983, 1986). Raatz et al. (1991, S. 51) kritisierten die betreffenden Testverfahren mit dem Argument, dass dem Testinhalt eine weniger wichtige Rolle zugesprochen werde als der Akzeptanz der Testverfahren:

„Meist wird das Konzept der inhaltlichen Validität zugrunde gelegt, wobei aber oft das Itemuniversum nicht genau definiert und auch nicht zu erkennen ist, nach welchem Prinzip die Aufgaben aus dem Universum ausgewählt werden. Trotz der Hinweise auf die Zustimmung von Experten oder auf die Übereinstimmung mit Lehrzielkatalogen oder Schulbüchern entsteht der Eindruck, dass die Augenscheingültigkeit eine größere Rolle als Inhaltsgültigkeit spielt“ (Raatz et al., 1991, S. 51).

Auf vielen Ebenen wurden Korrelationen zwischen verschiedenen Aufgabenteilen beobachtet (vgl. einen Überblick bei Klein-Braley, 1981; Vollmer, 1982), die laut vorherrschender Theorie eigentlich unabhängig voneinander sein sollten. Die faktorenanalytischen Untersuchungen zeigten, dass verschiedene Aufgabenteile (Untertests in einer Testbatterie oder in einem Testset) auf einem einzigen Faktor lagen, was nach Meinung der Linguisten die strukturalistische Sprachtheorie in Frage stellte. Die Sprachkompetenz, die nach den testtheoretischen Ergebnissen ein einheitliches Konstrukt bildete, forderte eine neue Sprachtheorie.

In den darauffolgenden Jahren (1983, 1986) beschrieb Oller Sprache als ein ganzheitliches Konstrukt bzw. Gebilde, das nicht künstlich in einzelne Elemente zerlegt werden könne. Sprache sollte ganzheitlich erfasst werden, sonst würde man ihre Natur verletzen und sie durch Tests nicht adäquat abbilden (vgl. Oller, 1973; Oller & Hinofotis, 1978; Oller, 1983, 1986). Süßmilch (1984a, b) stellte die Forderung an Sprachtests, den funktionalen Gebrauch der Sprache messen zu sollen (Performanzebene) und nicht nur die Kenntnis sprachlicher Regeln (z. B. Orthografie).

Der psycholinguistisch-soziolinguistische Ansatz

Der psycholinguistisch-soziolinguistische Ansatz zeichnet sich dadurch aus, dass die Globalbeherrschung der Sprache im Mittelpunkt steht (dazu Spolsky, 1973; Oller, 1973). Alle Aspekte der Sprache sollen gleichzeitig getestet werden. Spolsky (1978, S. viii) betont u.a. den notwendigen Bezug zur sprachlich-kommunikativen Seite und zur Sprachverwendung in alltagsrelevanten, realen Situationen: „the communicative competence [...] insists on the need to add a strong functional dimension to language testing“.

Von den Testverfahren des psycholinguistisch-soziolinguistischen Ansatzes wurde, wie von den Tests der psychometrisch-strukturalistischen Phase, erwartet, dass sie allen Gütekriterien entsprechen sollten (vgl. Klein-Braley, 1981, 1985a, b; Wockenfuß, 2009, S. 91). Den oben genannten Kriterien entsprach jedoch nur das Testformat Diktat (dazu Oller, 1972). Die Entwicklung von Sprachtests prägte in dieser Phase Spolskys Annahme, dass Sprache redundant und kreativ sei. Daher suchte man nach geeigneten Sprachtests, die dieser Theorie entsprachen. Zu dieser Zeit wurden Sprachtests wie das *Partial Dictation*¹⁴ (Johansson, 1973), der *Cloze-Test* (dazu Oller & Nevin, 1971), der *Multiple-Choice-Cloze-Test*¹⁵ (dazu Jonz, 1976) und der *C-Test* (Raatz & Klein-Braley, 1983) entwickelt und in der Praxis erprobt¹⁶. Die Testverfahren Cloze-Tests und C-Tests werden ausführlich in Kap. 4.3.1 und 4.3.2 beschrieben.

Zur Messung von Sprachleistungen in der Mutter-, Fremd- oder Zweitsprache werden heutzutage Testverfahren aus allen drei Ansätzen verwendet. Welcher Sprachtest eingesetzt wird, hängt letztlich von der jeweiligen Fragestellung ab. Wichtig sind einerseits das Ziel der Testung, die dem diagnostischen Verfahren zugrunde gelegt wird und andererseits das ausgewählte Sprachmodell (vgl. Kap. 1.3) bzw. der theoretische Ansatz (vgl. Kap. 3.1), der die Sprachkompetenz, die abgebildet werden soll, beschreibt. Den Testverfahren, die im Rahmen der vorliegenden Studie zum Einsatz kommen, liegen der psychometrisch-

¹⁴ Das *Partial Dictation* ist ein Teildiktat, bei dem ein Proband einen ihm vorliegenden beschädigten Text — d. h. Wörter oder Phrasen fehlen — während des Diktats ergänzen muss.

¹⁵ Im *Multiple-Choice Cloze-Test* muss der Proband aus verschiedenen Antwortmöglichkeiten das Wort herausfinden, welches sinnvoll in die Lücke passt.

¹⁶ Die genannten Testverfahren sind bei Klein-Braley (1997) vergleichend untersucht worden.

strukturalistische und der psycholinguistisch-soziolinguistische Ansatz zugrunde. Der Einsatz von C-Tests lässt sich durch den psycholinguistisch-soziolinguistischen Ansatz begründen, da C-Tests die Sprache ganzheitlich erfassen, ohne Zerlegung in einzelne Teildimensionen der Sprachkompetenz. Die mit den C-Tests kombinierten Lesetests erfassen jedoch nur bestimmte Dimensionen der Lesekompetenz und werden mithilfe des psychometrisch-strukturalistischen Ansatzes interpretiert.

3.2 Sprachtests und ihre Anwendungsbereiche

Im Rahmen der Globalisierung spielen (Fremd-)Sprachkenntnisse eine immer größere Rolle. Dadurch gewinnt auch die sprachenbezogene Diagnostik an Bedeutung. Mithilfe von Sprachtests lassen sich große Gruppen von Lernern testen und anhand der Ergebnisse in geeignete Sprachkurse oder spezielle Fördersprachkurse einteilen. Diese Einsatzmöglichkeiten sind auf verschiedene Zielgruppen (Muttersprachler, Fremdsprachen- und Zweitsprachenlerner, aber auch Kinder, Jugendliche und Erwachsene) und verschiedene Institutionen übertragbar, sowohl im schulischen Bereich als auch im Ausbildungsbereich und sogar im Berufsleben (z. B. Eignungstests bei Piloten oder Dolmetschern).

Das Ziel von Sprachtests im genannten Kontext ist es, den Lernern eine Rückmeldung über ihre sprachlichen Leistungen zu geben (Raatz & Klein-Braley, 1983). Darüber hinaus gibt es eine Vielzahl von Studien, die sich mit den individuellen Unterschieden beim Sprachenlehren und -lernen beschäftigen (z. B. im fremdsprachlichen Bereich Arbeiten von Dörnyei & Schmidt, 2001; Finkbeiner & Schnaitmann, 2001; Börner & Vogel, 2004). Diese sprachdiagnostischen Verfahren werden in verschiedenen Fachrichtungen als Individualtests zur Differenzialdiagnostik entwickelt und eingesetzt – zum Beispiel in der Logopädie, in der Entwicklungsphysiologie oder in der Sonderpädagogik. Diese Sprachtests sind als Einzeltests konzipiert, um individuelle Auffälligkeiten in sprachlichen Bereichen zu überprüfen (z. B. frühkindliche Störungen in der Sprachentwicklung) und eine entsprechende Therapie bzw. Förderung anbieten zu können (dazu Heinemann & Höpfner, 1993; Grimm, 1999; Wockenfuß, 2009,

S. 101). Im Rahmen der psychologischen Forschung werden ebenfalls sowohl individuelle als auch sprachdiagnostische Gruppenverfahren eingesetzt, z. B. bei der kindlichen Sprachentwicklung (dazu Trautner, 1991; Ehlich, 1996, 2007; Szagun, 2006).

In der empirischen linguistischen Forschung geht es mehr um die Überprüfung linguistischer Theorien, die „in dem jeweiligen diagnostischen Verfahren [...] eine befriedigende Operationalisierung findet“ (Raatz & Klein-Braley, 1983, S. 107). Bekannt ist in diesem Zusammenhang Spolsky, der sich mit der Sprachdiagnostik beschäftigt hatte. Spolsky (1981) unterteilt Sprachtests in zwei Arten: 1) strukturalistische Tests und 2) integrative Tests. Dieser Einteilung werden die oben beschriebenen Ansätze zur Konstruktion von Tests zugrunde gelegt.

- 1) In *strukturalistischen Sprachtests* wird Sprache in ihren einzelnen Komponenten oder Dimensionen getestet. Entweder wird sie in vier Fertigkeiten zerlegt (Hören, Sprechen, Lesen, Schreiben) oder in andere spezifische Sprachbereiche (Wortschatz, Grammatik, Aussprache etc.). Zu den strukturalistischen Sprachtests gehören Lesetests, die im Rahmen dieser Studie verwendet werden.
- 2) *Integrative Sprachtests* sind Verfahren, bei deren Bearbeitung gleichzeitig mehrere Fähigkeiten und Fertigkeiten eingesetzt werden müssen. Sprache wird hierbei ganzheitlich erfasst. Zu dieser Kategorie werden die C-Tests gezählt. C-Tests sind integrative schriftliche Sprachtests, mit denen die allgemeine Sprachbeherrschung erfasst werden kann. Auf dieses Testformat wird ausführlich in Kap. 4 eingegangen.

Sprachtests können u. a. mit verschiedenen Typen von Entscheidungen zusammenhängen (vgl. Grotjahn, 2000, S. 72) und werden im Folgenden nach Testfunktionen unterschieden.

1. *Einstufung* - hier geht es um die Auswahl der Sprachkurse, die gegenüber dem Leistungsstand der Probanden angemessen erscheinen.
2. *Selektion* - hier werden anhand der erbrachten Leistungen studien- oder berufsrelevante Entscheidungen getroffen.
3. *Evaluation* - als Grundlage für Unterrichtsmaßnahmen, zur Notengebung, zur Bewertung der Effizienz von Unterricht oder auch zur Selbsteinschätzung.

Sprachtests werden häufig in Bezug auf ihre Funktionen und ihre Mischformen unterschieden (Krekeler, 2005, S. 13f.).

a) Sprachtests mit Bezug zu einem Curriculum:

Lernfortschritttests (*progress tests, achievement tests*)

Kursabschlusstests (*exit tests, achievement tests*)

b) Sprachtests mit Bezug zu einem sprachlichen Konstrukt:

Feststellungsprüfungen (*proficiency tests*)

c) Sprachtests mit Bezug zur Zukunft:

Eignungstests (*aptitude tests*)

Einstufungstests (*placement tests*)

Zulassungsprüfungen (*selection tests* bzw. *gatekeeping tests*)

Diagnosetests (*diagnostic tests*)

Die Ergebnisse der Sprachtests mit curricularem Bezug werden vor allem mit Blick auf einen bestimmten Lernstoff interpretiert. *Lernfortschritts- und Kursabschlusstests* beziehen sich auf einen Lernprozess. Sprachtests mit Bezug zu einem sprachlichen Konstrukt geben Informationen über den Sprachstand der Kandidaten. Die Aufgabe von *Feststellungsprüfungen* ist die Erfassung von Sprachständen mit Bezug auf einen Maßstab (z. B. GER). Häufig werden die Ergebnisse mit dem Ziel interpretiert, Aussagen über die Angemessenheit der Sprachkenntnisse für eine bestimmte Aufgabe zu gewinnen (z. B. ausreichende Kenntnisse, um einen bestimmten Kurs in einem bestimmten Fach zu belegen).

Sprachtests, die sich auf die Zukunft beziehen, sollen zum Beispiel Erfolgsaussichten vorhersagen. In *Eignungstests* wird überprüft, ob eine Person über die notwendigen Voraussetzungen verfügt, um den Anforderungen eines

Bildungswegs zu genügen (z. B. Eignung für Piloten, für ein Medizinstudium). *Einstufungstests* verfolgen das konkrete Ziel der Kurszuweisung. Mithilfe von Einstufungstests werden Zuordnungen zu bestimmten Lerngruppen vorgenommen. Der Testinhalt orientiert sich am Inhalt der jeweiligen Kurse. *Zulassungsprüfungen* werden zur Überprüfung von Eingangsvoraussetzungen für ein bestimmtes Studium oder ein bestimmtes Fach verwendet.

Mit *Diagnostetests* können sprachliche Fähigkeiten der Probanden und sprachliche Defizite eingeschätzt werden. Mit diesen Informationen werden lernprozessrelevante Empfehlungen ausgesprochen. Für diese Zwecke können z. B. auch C-Tests genutzt werden (dazu ausführlich in Kap. 4).

Die Diskussion über Sprachtests wird im Bereich der Fremdsprachenforschung seit den sechziger Jahren des vergangenen Jahrhunderts intensiv und kontinuierlich geführt. Seit den achtziger Jahren begann eine Beschäftigung mit Tests für den Bereich Deutsch als Zweitsprache. Bei Sprachtests in der Muttersprache Deutsch lag der Schwerpunkt auf der Entwicklung von Sprachtests in der Grundschule – insbesondere, um Sprachfähigkeiten beim Eintritt in die Grundschule im Sinne einer Schulreife festzustellen und Schulleistungen beim Übergang in weiterführende Schulen objektiv zu beurteilen. Im Rahmen der PISA- und DESI-Studien (vgl. Kap 1.6.1 und Kap. 1.6.2) wurden die Sprachfähigkeit deutscher Muttersprachler im Alter von 15 Jahren in ihrer Muttersprache näher betrachtet und international miteinander verglichen. Neu ist die Diskussion um die Testung der sprachlichen Leistungen von *erwachsenen* Muttersprachlern, wie sie im *Sprachkompetenzprojekt Diagnose & Förderung* an der Universität Duisburg-Essen im Mittelpunkt steht.

In den folgenden Kapiteln wird ein Überblick über Sprachtests für verschiedene Altersstufen zusammengestellt, mit Blick auf die für Studierende relevanten Testformate. Ziel der ausführlichen Darstellung ist eine Suche nach Testverfahren, mit denen sich die allgemeine Sprachkompetenz unter besonderer Berücksichtigung von Lesefähigkeiten abbilden lässt. Hintergrund ist, dass in der Praxis häufig die Meinung vertreten wird, die Entwicklung neuer

Testverfahren sei unnötig, weil mittlerweile eine sehr große Auswahl an Tests zur Verfügung stehe, die problemlos auf andere Gruppen übertragbar seien. Im Folgenden werden nicht nur die verwendeten Testformate und ihre Ziele dargestellt, sondern auch einige Kritikpunkte beleuchtet, falls diese in der Testpraxis bekannt sind.

3.3 Sprachtests in der Schule

Sprachliche Bildung und Förderung ist ein Schwerpunkt in allen Bildungsplänen, denn je besser die Entwicklung der Sprachkompetenz gelingt, desto günstiger sind die Voraussetzungen für den Erfolg in der Schule und zur Teilhabe am gesellschaftlichen Leben (vgl. den Kompetenzbegriff von PISA)¹⁷. Spätestens wenn Kinder eingeschult werden, werden sie mit den Sprachbereichen Lesen und Schreiben systematisch konfrontiert. Ein großer Teil der gegenwärtigen Forschung nimmt die ersten vier Jahre der schulischen Ausbildung in den Blick, wie schon oben erwähnt. Diese Testverfahren und Testformate sind zum Abgleich mit meiner Adressatengruppe (Studierende) von geringerer Bedeutung. Daher gehe ich im Folgenden exemplarisch auf einzelne Testverfahren ein, die für die Sek. I und Sek. II entwickelt wurden.

Ein wichtiges Verfahren zur Sprachstandsdiagnose ist die *Hamburger Schreibprobe* (HSP 1-9). Sie ist nicht nur auf die Grundschule beschränkt, sondern wird auch bei älteren Schulkindern eingesetzt (vgl. May, 1996, 1999; May & Malitzky, 1999; May, Malitzky & Wieluf, 2001). Die HSP ist ein Rechtschreibtest, der dazu dient, das Rechtschreibkönnen und die grundlegenden Rechtschreibstrategien von Schülern einzuschätzen. Die HSP kann als Gruppen- und Einzeltest für alle Schulformen eingesetzt werden.

Vom Lehrer werden dabei Testwörter bzw. Testsätze vorgelesen, die in den Testheften anhand von Illustrationen veranschaulicht werden. Für die Bestimmung des Lernstandes im Primarbereich stehen die Versionen HSP 1-4/5

¹⁷ Hier wird nicht auf die Sprachtests im Elementarbereich eingegangen. Einen umfangreichen Überblick über die Sprachtests für das Vorschulalter in verschiedenen Bundesländern gibt es bei Roth (2008) und Lisker (2010)¹⁷.

zur Verfügung. Die Aufgabe der Schüler besteht darin, Einzelwörter und Sätze zu identifizieren und ins Heft zu schreiben. Ihre Bedeutung wird zusätzlich durch Bilder veranschaulicht. Die Schüler können in ihrem eigenen Tempo schreiben, nachdem sie mit den Bild-Wörtern vertraut gemacht worden sind. Die Sätze werden dann vom Lehrer diktiert, wenn die Schüler mit der Bild-Wort-Aufgabe fertig sind¹⁸.

Die Auswertung erfolgt auf Wort- und Graphemebene: Neben der Zahl der richtig geschriebenen Wörter wird auch die Anzahl der richtig geschriebenen Grapheme ermittelt. Durch die Bestimmung von Strategieprofilen (alphabetische, orthografische, morphematische, wortübergreifende Strategien) werden Störungen im Aneignungsprozess differenziert erfasst.

Mit der HSP 5-9 werden orthografische Grundfertigkeiten erfasst, über die jeder kompetente Schreiber im Deutschen verfügen muss, um die meisten Wörter und Sätze richtig schreiben zu können. Die HSP 5-9 umfasst eine Aufgabe mit einer Textvorlage, in der teilweise falsch geschriebene Wörter vorkommen, die korrigiert werden müssen. Außerdem sollen Satzzeichen ergänzt werden¹⁹.

Ein solcher Test wäre für die Gruppe der Studierenden ungeeignet, da sie bereits über schriftsprachliche Fähigkeiten verfügen, die über orthografische Grundfähigkeiten hinausgehen und die Strategien, die beim Schreiben getestet werden, hier nicht zur Geltung kommen.

Die Studie *Hamlet 3-4* ist ein aus zwei Teilen bestehender Test, der das erreichte Leseverständnis in der 3. und 4. Klasse überprüft. Mit einem Worterkennungs-Test (40 Wortzuordnungen zu je 4 Bildern) werden grundlegende Informationen zur Lesefertigkeit und -geschwindigkeit erhoben. Die Ergebnisse aus einem Leseverständnistest, der sich aus 10 Texten (Sach-, Gebrauchs- und Erzähltexte) mit Fragen im Multiple-Choice-Verfahren zusammensetzt, erlauben eine Zuordnung der Schüler zu einer bereits erreichten Stufe des sinnverstehenden,

¹⁸ HSP 1-4/5-Aufgaben sind unter folgendem Link nachvollziehbar: http://foerderung.bildung-rp.de/fileadmin/user_upload/sonderpaedagogik.bildung-rp.de/PL_ehemPZDateien/7374.d4-6.pdf, zuletzt gesehen am 27.05.2013.

¹⁹ HSP-Manual kann unter folgender Adresse heruntergeladen werden: http://www.vpm-verlag.de/tl_files/Katalog_Upload/Grundschule/Diagnose/hsp/010096_S5-59_Leseprobe.pdf; zuletzt gesehen am 27.05.2013.

stillen Lesens (vgl. Lehmann, Peek & Poerschke, 1997, S. 5). In der Hamlet-Studie wurden u. a. hohe Korrelationen zwischen den Schülerleistungen und der Einschätzungen durch die Lehrkräfte festgestellt²⁰.

Die zu messenden Teilkompetenzen (Worterkennung und Lesegeschwindigkeit, sinnentnehmenden Lesen) sind bei Studierenden in dieser Form weit entwickelt. Wollte man den Test für die vorliegende Studie übernehmen, könnte man nur den Leseverständnistest (LV-Test), der vorher an das Niveau der Studierenden angepasst werden müsste, verwenden. Im Unterschied zum dargestellten LV-Test müsste man für Studierende andere Grundlagentexte auswählen. Denn auf der Grundlage alltagsrelevanter Texte lässt sich keine Leistungsstreuung in der Gruppe erzielen, da die Antworten zum Text mit dem Vorwissen der Probanden zu lösen wären. Daher ist der Einsatz auch dieses Testformats bei Studierenden nicht sinnvoll.

Für die Erfassung der Lesefähigkeit, auch im Hinblick auf Textverständnis, wurde in Deutschland die *IGLU-Studie*²¹ durchgeführt (IGLU 2004, IGLU-E 2006), die Schulkinder in der 4. Jahrgangsstufe testet. IGLU erfasst diese Kernkompetenz bei Schülern am Ende ihrer Grundschulzeit und richtet den Fokus nicht nur auf die Untersuchung der Schülerleistungen, sondern auch auf deren schulische und familiäre Bedingungsfaktoren. Die eingesetzten Testaufgaben weisen unterschiedliche Schwierigkeitsgrade auf und enthalten verschiedene Textsorten, zum Beispiel literarische Texte (z. B. Kurzgeschichten) und informierende Texte (z. B. altersgerechte Lexikonartikel). Mit Hilfe von Fragebögen wird bei IGLU (vgl. Bos et al. 2006) zudem erfasst, wie Kinder über das Lesen denken und wie gern und häufig sie lesen²². Kinder mit einem erhöhten Förderbedarf werden durch *textimmanente* (Abfrage der im Text explizit genannten Informationen) und *wissensbasierte* Verstehensleistungen (Verknüpfung mit eigenem Wissen) identifiziert (vgl. dazu auch das Kapitel 1.5 und 1.6.1 zur Überprüfung der Sprachkompetenz). Zur Aufdeckung der mit den

²⁰ Informationen zu Korrelationen zwischen dem Hamlet-Test und den anderen Tests sind unter folgender Adresse zu finden: www.testzentrale.de/.../hamburger-lesetest-fur-3-und-4-klassen.html; zuletzt gesehen am 27.05.2013.

²¹ IGLU - Internationale Grundschul-Lese-Untersuchung.

²² <http://www.bmbf.de/de/6626.php>.

Leistungen der Schüler zusammenhängenden Faktoren werden äußere Rahmenbedingungen (z. B. Schulcurricula, Richtlinien zur Lehrerbildung), sozialräumliche Bedingungen der Schule, familiäres Umfeld usw. dokumentiert. Lesekompetenz wird über vier Verstehensprozesse erfasst (vgl. Bos et al. 2011):

1. Erkennen und Wiedergeben explizit angegebener Informationen in einem Text,
2. einfache Schlussfolgerungen,
3. komplexe Schlussfolgerungen ziehen und begründen, Interpretieren des Gelesenen und
4. Prüfen und Bewerten von Inhalt und Sprache.

Die Studie sorgte für Diskussionsstoff in Bezug auf die Schullaufbahnpfehlungen der Lehrkräfte. Es wurde deutlich, dass die Chance eines Kindes aus einer Familie mit sozial-ökonomisch höherem Status, eine Gymnasialempfehlung zu erhalten, selbst bei gleichen kognitiven Grundfähigkeiten und gleicher Lesekompetenz höher ist als für eines aus einem Haushalt mit niedrigerem sozial-ökonomischen Status.

Eine weitere internationale Studie zur Erfassung der Lesekompetenz trägt den Namen PIRLS (vgl. Bos et al. 2001). Es handelt sich eher um eine Erweiterungsuntersuchung von IGLU, die einen internationalen Vergleich ermöglicht, für die aber noch keine Ergebnisse vorliegen. Die PIRLS-Studie hat eine ähnliche Struktur wie die IGLU-Studie.

In Bezug auf die in der IGLU-Studie dargestellten Aufgabenformate kann festgehalten werden, dass sich die Lesefähigkeiten der Studierenden mit diesen Aufgaben abbilden lassen. Die Teilkompetenz Lesen lässt sich über die vier von Bos et al. (2011) beschriebenen Verstehensprozesse beschreiben (s.o.). Allerdings sollten für die Gruppe der Studierenden andere Textsorten für den Leseimpuls ausgewählt werden. Mit dem Einsatz literarischer Texte würde man jene Studierenden benachteiligen, die andere Fächer als Deutsch studieren. Somit hätte man das Gütekriterium *Fairness* verletzt. Die Verwendung nur von informierenden Texten auf der Grundlage eines Lexikonartikels könnte aus den oben skizzierten Gründen jedoch ebenfalls problematisch sein.

Mit dem Projekt VERA (Vergleichsarbeiten der 3. und 8. Klassen) soll mittels Testverfahren der schulfachliche Kompetenzstand der Schüler einer Klasse

objektiv und zuverlässig überprüft werden. Ziel von VERA ist die Kontrolle der Unterrichtsentwicklung mit Bezug auf Leistungsmessung von Schülern, um festzustellen, ob und inwiefern die deutschlandweit festgelegten Standards eingehalten werden. Die Vergleichsarbeiten (VERA 3) orientieren sich an Anforderungen der länderübergreifend verbindlichen Bildungsstandards für den Primarbereich und untersuchen die Bereiche Lesen, Schreiben und Mathematik. Für die 3. Klassen gliedern sich die Aufgaben im Fach Deutsch (das hier im Mittelpunkt steht) in *Leseaufgaben*, *Schreibaufgaben*, *Sprachgebrauchsaufgaben* und *Zuhöraufgaben*. Der eigentliche Wert der Lernstandserhebungen liegt in den kompetenzorientierten Rückmeldungen auf Basis inhaltlich beschriebener Fähigkeitsniveaus für die einzelnen Klassen, deren Ergebnisse durch den Vergleich untereinander und mit den Landeswerten aufschlussreiche Informationen über Stärken und Schwächen der Schüler geben. Mit Berücksichtigung des sozialen Kontextes können sich Klassen auch mit einer Gruppe von Klassen vergleichen, deren Schülerschaft der eigenen Klasse ähnlich ist (vgl. Helmke & Hosenfeld, 2003; Bremerich-Vos et al. 2005; Thonke et al. 2008; Groß Ophoff et al. 2008; Isaac, 2008). Das Testverfahren wurde dennoch stark kritisiert, was Objektivität und Interpretation der Testergebnisse angeht (vgl. die unten aufführende Kritik dazu).

Spezielle Aufgaben gibt es mittlerweile nicht nur für die 3. Klasse, sondern auch für die 8. Klasse für die Fächer Deutsch, Englisch, Mathematik und Französisch. Im Folgenden wird auf einzelne Aufgaben (-formate) im Fach Deutsch eingegangen. Für die 8. Klassen lassen die Aufgaben sich in ähnliche Typen einteilen; allerdings werden sie durch einen weiteren Aufgabentyp vervollständigt, nämlich *Orthografieaufgaben*. Genutzt werden dabei kontinuierliche und teilweise diskontinuierliche Texte²³.

Bei *Leseaufgaben* werden den Schülern verschiedene Textsorten vorgelegt (z. B. Sachtexte und literarische Texte), zu denen Multiple-Choice-Fragen gestellt werden. Zum Teil gibt es auch halboffene Aufgaben, die den Schülern die Möglichkeit bieten, die Antwort selbst zu formulieren.

²³ Beispielaufgaben sind auf der Seite des IQB zu finden: <https://www.iqb.hu-berlin.de/vera/aufgaben>, zuletzt gesehen am 16.06.2013.

In den vorgegebenen *Schreibaufgaben* sollen Schüler ein Inserat bzw. einen Antrag selbstständig formulieren, eine Zeitungsnachricht schreiben und eine argumentierte Stellungnahme zu einem Leserbrief erstellen.

In *Sprachgebrauchsaufgaben* werden verschiedene Bereiche überprüft, z. B. Pronomenverwendung, Groß- und Kleinschreibung oder Kenntnisse über die Wortarten.

Die Aufgabenstellung der *Zuhöraufgaben* besteht darin, einen Textausschnitt (z. B. Ausschnitt aus einem Hörbuch, Hörspiel oder Radiobeitrag) zu hören und gleichzeitig Multiple-Choice- oder Richtig/Falsch-Aufgaben zu lösen.

Bei *Orthografieaufgaben* sollen z. B. einzelne Fehler erkannt und korrigiert werden oder aufgrund des vorhandenen Regelwissens soll die Rechtschreibung begründet oder unbekannte Wörter im Lexikon nachgeschlagen werden.

Die VERA-Studie wurde stark kritisiert. Nachfolgend wird auf wesentliche Kritikpunkte eingegangen, die vom Grundschulverband²⁴ formuliert wurden.

- Durch die von VERA gewählten Aufgabenformate würden wesentliche Bereiche und Kompetenzen der Bildungsstandards nicht erfasst (zum Beispiel beim Schreiben: Texte planen und überarbeiten; beim Lesen: Texte begründet für das Leseziel auswählen). Die Aufgabenformate ließen zudem auch keine Beobachtung der angewendeten Strategien zu, die Kinder bei der Lösung der Aufgaben einsetzen.
- Die Auswertungsvorgaben ließen keine alternativen Lösungen oder vom Lösungsschlüssel abweichende Antworten zu. Auch wenn die Lösungen zeigten, dass Kinder im Sprachgebrauch kreativ sind, würden die Lösungen als falsch bewertet.
- Testergebnisse aus der VERA-Studie ließen keine objektiven und eindeutigen Schlussfolgerungen zum Lernstand einzelner Kinder zu, da viele Einflussfaktoren wie z. B. Tagesleistung eines Kindes nicht berücksichtigt würden. Die Ergebnisse würden nicht nur an Lehrkräfte weitergeleitet, sondern auch in der Öffentlichkeit benutzt.
- Testergebnisse würden dadurch verfälscht, dass viele Lehrkräfte u. a. die Arbeiten im Vorfeld üben und den Kindern bei der Testdurchführung helfen.

²⁴ Pressemitteilung vom 11.05.2009: VERA 3 – die überschätzten Vergleichsarbeiten. Kritik und Alternativen, unter:
http://www.grundschulverband.de/fileadmin/Bildungspolitik/3_VERA_Kritik_Alternativen.09.pdf,
zuletzt gesehen am 22. 05.2013.

Auf die vom Grundschulverband geäußerte Kritik folgte eine Pressemitteilung²⁵, in der folgende Punkte zusammengefasst wurden.

- Lernstandserhebungen ermöglichen eine schulübergreifende Standortbestimmung. Die Schulen können sich mit den Ergebnissen vergleichen, die in Nordrhein-Westfalen insgesamt und an Schulen mit ähnlichen Standortvoraussetzungen erreicht wurden. Eine solche schulübergreifende Einordnung hilft, den Erfolg der pädagogischen Arbeit besser einzuschätzen.
- Mit der Berücksichtigung des schulischen Einzugsgebietes (sozialer Kontext) können sich Klassen auch mit einer Gruppe von Klassen vergleichen, deren Schülerschaft der eigenen Klasse ähnlich ist.
- Bei der Auswertung der Ergebnisse erfassen die Lehrkräfte auch den Sprachhintergrund der Kinder. Unter "Deutsch nicht dominant" werden alle Schülerinnen und Schüler zusammengefasst, bei denen – unabhängig von Nationalität und Geburtsort – im Alltag in der Regel nicht Deutsch gesprochen wird.
- Vergleichsarbeiten stellen eine wichtige Grundlage für eine systematische Schul- und Unterrichtsentwicklung dar. Sie bieten den Lehrkräften zuverlässige Informationen, welche Lernergebnisse die Schülerinnen und Schüler erreicht haben und inwieweit die fachlichen Anforderungen der Lehrpläne und der nationalen Bildungsstandards erfüllt wurden.

Bei VERA-8 (2009) wurden einige Änderungen vorgenommen, die z. B. die Klassen nach Kontextgruppen (Klassen mit günstigen und weniger günstigen Rahmenbedingungen) unterscheiden und weitere wichtige Variablen (z. B. Sprachhintergrund der Kinder bzw. Eltern) erfassen. Dadurch sollte die Standortbestimmung und eine damit verbundene realistische Einschätzung der Testergebnisse gesichert werden.

Analysiert man die Aufgaben, die für VERA 3 und 8 entwickelt wurden, erscheint es fraglich, ob diese auch im universitären Bereich angewendet werden könnten (z. B. Zuhöraufgaben). Auch die Einführungsseminare, in denen an der Universität getestet wird, erfüllen nicht die Testbedingungen, die für eine objektive Datenerhebung mit dem Instrumentarium von VERA erforderlich wären. Zwar könnte man die für VERA entwickelten Orthografieaufgaben mit Schreibaufgaben kombinieren, die dann in Bezug auf Sprachrichtigkeit und

²⁵ Ministerium für Schule und Weiterbildung. Ergebnisse der Vergleichsarbeiten (VERA), Klasse 3, 2009, vom 4. 10.2009.

Argumentation bzw. den Inhalt bewertet würden. Allerdings ist im Rahmen der vorliegenden Studie der Einsatz von Schreibaufgaben bei einem Screening, an dem große Gruppen von Studierenden teilnehmen, aus zeitökonomischen Gründen nicht möglich²⁶.

Wollte man das Format der Leseaufgaben von VERA übernehmen, so müssten sie dem Niveau und dem Wissen der Studierenden angepasst werden; dann sollte man z. B. als Textgrundlagen pädagogisch-propädeutische Texte nutzen, weil sie einerseits für die Studierenden thematisch mehr relevant sind und andererseits sprachlich anspruchsvoller.²⁷

Eine weitere wichtige Studie, die in Deutschland durchgeführt wurde, ist die *DESI-Studie* (Deutsch Englisch Schülerleistungen International)²⁸. Das Ziel von DESI ist die Erfassung sprachlicher Kompetenzen in den Fächern Deutsch und Englisch in der 9. Klasse. DESI ist die erste Schulleistungsstudie, für die auf nationaler Ebene in Deutschland fachbezogene Tests entwickelt wurden. In der Studie wurden erstmals rezeptive und produktive Kompetenzen im Deutschen und Englischen untersucht (vgl. Beck & Klieme 2007; DESI-Konsortium 2008).

Im DESI-Projekt, in dem Kompetenzen von Schülern der 9. Jahrgangsstufe erfasst und in ihrer Entwicklung analysiert werden, wird Sprachkompetenz als Gesamtkompetenz verstanden, die in spezifische Teilkompetenzen aufzuschlüsseln ist (vgl. dazu Kap. 1 zur Erfassung von Sprachkompetenz). Einerseits sollen linguistische Komponenten (Morphologie, Syntax, Lexik) überprüft, andererseits Strukturen analysiert werden, die produktive und rezeptive Teilkompetenzen verbinden (vgl. Jude & Klieme 2007, S. 10). Beides erfolgt mit dem Ziel, spezifische sprachliche Fähigkeiten individuell zu diagnostizieren und ggf. zu fördern. Das Gesamtmodell von DESI beschreibt die Teilbereiche, die mit unterschiedlichen Testinstrumenten zu einem ganzheitlichen Sprachbild führen. Diese sind jedoch nicht als unabhängig

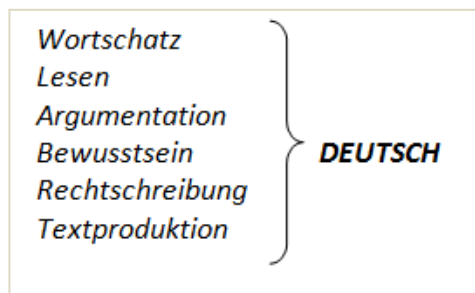
²⁶ Mit der Erfassung von schriftsprachlichen Fähigkeiten der Studierenden beschäftigt sich das Sprachkompetenzprojekt Diagnose & Förderung an der Universität Duisburg-Essen (vgl. Scholten-Akoun & Baur, 2012; Scholten-Akoun, Tischmeyer & Mashkovskaya, 2013; Bremerich-Vos et al., 2013).

²⁷ Vgl. dazu die oben genannte Literatur.

²⁸ Dazu auch das Kap. 1.6.2.

voneinander zu betrachten, sondern bilden im Sinne der sprachlichen Mehrdimensionalität eine Basis für die übergeordnete Sprachkompetenz (vgl. Gesamtmodell DESI-Kompetenzen, Jude & Klieme 2007, S. 18). Für die Sprachkompetenz im Deutschen wären das folgende Komponenten:

Abb. 5: Gesamtmodell DESI-Kompetenzen



Quelle: Jude & Klieme 2007, S. 18

Dieser Aufteilung der Kompetenz in Teilkompetenzen entsprechend kamen in DESI sechs Aufgabengruppen zur Messung sprachlicher Kompetenzen zum Einsatz. Lesekompetenz, Textproduktion (Schreiben von Briefen) und Sprachbewusstheit wurden sowohl zu Beginn als auch am Ende des neunten Schuljahres getestet. Am Ende des Schuljahres wurden dieselben Tests wiederholt und drei zusätzliche Tests durchgeführt: Rechtschreiben, Wortschatz und Argumentation. Der von DESI verwendete Lesetest kombiniert literarische Texte und Sachtexte²⁹ und umfasst 55 Fragen. Das theoretische Grundverständnis des Lesekompetenzmodells von DESI stellt drei Aspekte in den Vordergrund:

- Die Benennung spezifischer Textstellen, auf die sich die Fragen oder die Leser beziehen;
- Die Aktivierung des als vorhanden vorausgesetzten und teilweise gelernten Wissens;
- das mentale Modell als Ziel des Textverstehens.

Es werden verschiedene Strategien (nicht nur Lesestrategien, sondern auch Anwendung der angebotenen Textsorte) abgefragt, die die Schüler der 9. Klasse

²⁹ Sachtexte enthalten im Unterschied zur PISA-Studie keine Grafiken und Tabellen, da diese Textsorte in den Lehrplänen nur eine geringe Bedeutung hat.

laut Lehrplänen der Länder beherrschen sollen (vgl. Nold & Willenberg 2007, S. 39).

Die Schreibkompetenz wurde in DESI über zwei Briefe erfasst: ein Beschwerdebrief und ein persönlicher Brief an einen Freund oder eine Freundin. Analysiert wurden inhaltliche und argumentative Aspekte beim Schreiben (z. B. Basisinformationen, Forderungen, Konsequenzen) sowie Textsortenangemessenheit (z. B. richtiger Aufbau, Stil, Wortwahl und formale Kriterien). Bei der Auswertung wurden zwei Teilkompetenzen differenziert:

- eine semantisch-pragmatische Teilkompetenz
- eine sprachsystematische Teilkompetenz

Es wurde eine Einstufung in vier verschiedene Kompetenzniveaus³⁰ vorgenommen, die sich an den textsortenspezifischen Aufgabenlösungen und Sprachsystematiken orientieren (vgl. Harsch et al. 2007; Hartig, 2007; Willenberg, 2007; Eichler, 2007).

Übergreifend im Gesamtmodell von DESI sind die Fertigkeiten Lesen und Schreiben. Andere sprachliche Aspekte wie Argumentation oder Strukturiertheit vervollständigen diese und sind somit Bestandteil dieser Teilbereiche³¹.

Betrachtet man die dargestellten DESI-Aufgaben in Bezug auf die Anwendung für Studierende, könnten die Leseaufgaben zur Erfassung der Teilkompetenz Lesen mit entsprechender Anpassung für die Untersuchungsgruppe übernommen werden, z. B. in Bezug auf die Textsorte (s. Begründung für VERA-Aufgaben). Da Schreibaufgaben, wie oben bereits gesagt, keine ökonomische Auswertung zulassen, werden sie für die vorliegende Studie als ungeeignet eingeschätzt. Stünde eine Gruppe geschulter Auswerter zur Verfügung stehen, die die Texte der Studierenden bewerten können, ließe sich auch die Schreibaufgabe in modifizierter Form übernehmen. Die angebotenen Schreibimpulse jedoch wären für die Adressatengruppe der Studierenden zu einfach. Hier wäre – wie bereits bei der Diskussion der Leseaufgaben bei VERA ausgeführt – eine andere Grundlagentextauswahl notwendig.

³⁰ Statt fünf Kompetenzniveaus, die in der PISA-Studie festgelegt wurden.

³¹ Ausführlich dazu bei Klieme et al. (2006): <http://www.dipf.de/de/projekte/pdf/biqua/desi-zentrale-befunde>, zuletzt gesehen am 27.05.2013.

Zusammenfassend lässt sich festhalten, dass insbesondere in den Studien von VERA und DESI Anregungen für veränderte Testkonstruktionen enthalten sind, die im Rahmen weiterer Vergleichsforschungen mit dem C-Test genutzt werden können.

3.4 Sprachtests für erwachsene Sprecher an der Universität

Unter erwachsenen Sprechern werden hier Jugendliche im Übergang von der Oberstufe zur Universität verstanden. Für diese Untersuchungsgruppe gibt es außer der Abiturprüfung keine Studien, die den Sprachstand dieser Kandidaten umfassend messen. Mit dem schulischen Abschluss und dem Bestehen der Abiturprüfung wird den potentiellen Studienbewerbern bescheinigt, dass sie über ausreichende studienrelevante sprachliche Fertigkeiten verfügen.

In den Anforderungen für das Zentralabitur werden mehrere sprachliche Aspekte erfasst, auf die nachfolgend zusammenfassend eingegangen wird. Der Prüfling soll laut den für das Fach Deutsch formulierten Anforderungen für die Darstellungsleistung³² (2012) folgende sprachliche Bereiche beherrschen:

- seinen Text kohärent, schlüssig, stringent und gedanklich klar strukturieren (*Argumentation, Strukturiertheit*)³³.
- unter Beachtung der fachsprachlichen und fachmethodischen Anforderungen formulieren (*Strukturiertheit*).
- Aussagen durch angemessenes und korrektes Zitieren belegen (*Textproduktion*).
- sich allgemeinsprachlich präzise, stilistisch sicher und begrifflich differenziert ausdrücken (*Bewusstheit*).
- syntaktisch sicher, variabel und komplex (und zugleich klar) formulieren (*Bewusstheit, Textproduktion, Wortschatz*).

³² <http://www.standardsicherung.schulministerium.nrw.de/abitur-gost/fach.php?fach=1>, zuletzt gesehen am 27.05.2013.

³³ Die in Klammern kursiv markierten Begriffe sind dem Gesamtmodell aus der DESI-Studie zuzuordnen und machen die Zuordnung der Aufgabentypen zu verschiedenen Anforderungen deutlich (vgl. Kap 1.6.2).

- korrekt unter Berücksichtigung standardsprachlicher Normen schreiben und standardsprachliche Normen berücksichtigen (*Textproduktion, Rechtschreibung, Bewusstheit*).

Diese Beschreibung der einzelnen Bereiche konzentriert sich sowohl auf die korrekte Verwendung von sprachlichen Strukturen als auch auf die kommunikative Angemessenheit, die als Ziel des kommunikativen Ansatzes zu verstehen ist.

Im universitären Bereich existieren bislang keine fächerübergreifenden Standards bezüglich der sprachlichen Fertigkeiten von Studierenden. Curriculare, fachrelevante und sprachliche Anforderungen an Studierende (Beispiele aus verschiedenen Universitäten)³⁴ sind nur in den sprachlichen Fächern zu finden und universitätsabhängig. Die Untersuchungen mit den Studierenden beruhen in der Regel auf Messungen einzelner sprachlicher Dimensionen, von denen auf Sprachkompetenz geschlossen wird (z. B. Messung von Schreibfertigkeiten).

Mehrere Studien berichten darüber, dass den Studierenden wissenschaftssprachliche Kompetenzen und basale Sprachfertigkeiten fehlen (vgl. Augst & Feilke, 1993; Becker-Mrotzek 1995, 1997, 2003; Pospiech 2005; Steinhoff, 2007; Schindler & Siebert-Ott, 2010). Betont werden Defizite in folgenden Bereichen: mangelhafte Rechtschreibung und Grundlagen der Syntax, basale Fertigkeit zum informationsentnehmenden Lesen und die Fähigkeit, gehörte und gelesene Informationen schriftlich sinnvoll zusammenzufassen (dazu Pospiech, 2005; Pohl, 2007, Bremerich-Vos et al., 2013). In diesem Zusammenhang wird darüber diskutiert, inwieweit Studienerfolg und Studienzeiten mit der Beherrschung studienrelevanter sprachlicher Fertigkeiten zusammenhängen. Wenn etwa ein Zusammenhang zwischen mangelhafter sprachlicher Basis und den Abbruchquoten an den Universitäten besteht, wird die Notwendigkeit von Sprachstandfeststellungstests für Studierende deutlich.

Vor allem zwei Ziele sollen durch das Studium an der Universität erreicht werden: Studenten sollen zum einen verständlich und zum anderen richtig

³⁴ Hier ist ein Beispiel von der Universität zu Köln, Anforderungen an die sprachlichen Leistungen von Studierenden im Fach Deutsch: <http://www.uni-koeln.de/phil-fak/deutsch/aktuelles/SprachlicheMaengel.pdf>, zuletzt gesehen am 23.02.2013; Universität Hildesheim (Institut für deutsche Sprache und Literatur): <http://www.uni-hildesheim.de/8770.pdf>; zuletzt gesehen am 23.02.2013.

schreiben. Schreiben bedeutet dabei, einen Inhalt in eine Form zu bringen, die eine erfolgreiche Kommunikation mit anderen ermöglicht. „Viele Beispiele von der Bedienungsanleitung bis zu Lehrbüchern zeigen aber, dass diese Kommunikation nicht optimal gelingt“, betont Hans Krings von der Universität Bremen³⁵. „Und das Bildungssystem trägt wenig dazu bei, die Fähigkeit zum verständlichen Schreiben auszubilden“³⁶. Weitere Beispiele verdeutlichen ähnliche Situationen wie an den anderen Universitäten.

Constantin Körner bemängelt in einem Artikel vom 23.08.2012 in *Legal Tribune Online*³⁷ die Sprachkompetenz von Jura-Studenten an der Ruhr-Universität Bochum. Die sprachlichen Schwächen seien insbesondere für Juristen eine „peinliche Angelegenheit“. Schließlich, so Mühl vom Lehrstuhl für Kriminologie, Kriminalpolitik und Polizeiwissenschaft der Universität Bochum, bemühen diese sich schließlich um die Anwendung von Recht und Gesetz. Doch wie soll dies gelingen, wenn sie schon die Regeln der Rechtschreibung und Grammatik nicht beherrschen? Die oben geschilderten Probleme und Mängel im Bereich der Lesekompetenz können Juristen direkt an ihrem Zugang zum Recht behindern.

Ähnlich schildert auch Dr. Andreas Belwe die Situation an der Dortmunder Universität³⁸. Die Schreib- und Lesekompetenz von Studenten wird zunehmend schlechter, berichtet der ISM-Dozent. Belwe schätzt, dass rund 60% der Studierenden mit Schwierigkeiten wie fehlerhafte Grammatik, Interpunktion, Wortwahl und mangelhafter Orthografie zu kämpfen haben. Schreiben wird nicht mehr als schöpferischer Prozess angesehen, bei dem sich neue Gedanken entwickeln, sondern vielmehr als lästige Aufgabe. Erschwerend kommt hinzu, dass vielen Studenten die notwendige Übung fehlt. Nach Belwe ist „das Netz

³⁵ 2007, aus Anlass der Jahrestagung (2007) der Gesellschaft für angewandte Linguistik in Hildesheim zum Thema „Schlüsselqualifikation Sprache“, forderte Hans Krings (Universität Bremen) an deutschen Hochschulen die Einrichtung von Schreibzentren nach amerikanischem Vorbild. Deutschland brauche für seine Wirtschaft nicht nur Ingenieure und Naturwissenschaftler, sondern auch Mitarbeiter, „die wirklich verständlich schreiben können“.

³⁶ „Schlüsselqualifikation Sprache. Nur jeder vierte Student kann verständlich schreiben“, vom 29.09.2007, unter: <http://www.faz.net/aktuell/gesellschaft/schlüsselqualifikation-sprache-nur-jeder-vierte-student-kann-verstaendlich-schreiben-1459956.html>, zuletzt gesehen am 23.02.2013.

³⁷ <http://www.lto.de/recht/studium-referendariat/s/pilotprojekt-lesen-bildet-mehr-sprachkompetenz-fuer-jurastudenten>, zuletzt gesehen am 23.02.2013.

³⁸ Dr. Andreas Belwe ist als Dozent für wissenschaftliches Arbeiten und Business Ethics an der International School of Management (ISM) in Dortmund tätig.

zweifelloos eine geniale Erfindung, das Vieles möglich macht. Allerdings muss es Werkzeug bleiben und darf nicht als Spielzeug das Leben des Einzelnen und der Gesellschaft dominieren. Alles soll möglichst schnell gehen, am besten auf einen Klick zu haben sein“, so Belwe, „doch das funktioniert beim Lesen und Schreiben nicht.“³⁹

Weitere Kritikpunkte kommen von Schulen, die Referendare für den Vorbereitungsdienst annehmen. Die Klagen betreffen u. a. die Schriftsprache von Lehramtsstudierenden, die nicht in der Lage seien, den E-Mail-Verkehr zwischen der Schule und den Eltern zu bewältigen (vgl. Bremerich-Vos et al., 2013).

Mehrere Projekte wurden mit dem Ziel initiiert⁴⁰, erstens die Sprachkompetenz von Studierenden zu erheben und zweitens geeignete (kostengünstige) Testverfahren zu entwickeln, um sprachliche Defizite zu diagnostizieren (z. B. bei der Einschreibung oder spätestens im Grundstudium) und Fördermaßnahmen im Rahmen des Studiums anbieten zu können.

3.5 Sprachtests für den Hochschulzugang ausländischer Studienbewerber

Sprachtests werden auch bei ausländischen Studienbewerbern als Zulassungsprüfungen oder Eignungstests eingesetzt (Kap. 2.2), um zu überprüfen, ob für das Studium notwendige sprachliche Fertigkeiten vorhanden sind. Erreichen die Probanden die Stufen C1 oder C2, können sie eine Hochschulberechtigung bekommen. Die Sprachtests stellen eine differenzierte Erhebung der Sprachteilkompetenzen (Hören, Lesen, Sprechen und Schreiben) dar, die mit verschiedenen Testaufgaben erfasst werden. Die Studierfähigkeit ausländischer Studierenden wird mit folgenden Testformaten überprüft: die

³⁹ Online-Zeitungsartikel „Verschlechterung der Schreib- und Lesekompetenz bei Jugendlichen und Studenten. ISM-Dozent erklärt Gründe und die Rolle von Medien, Eltern, Schulen und Universitäten“, vom 15.02.2012, <http://www.presseportal.de/pm/70776/2199120/verschlechterung-der-schreib-und-lesekompetenz-bei-jugendlichen-und-studenten-ism-dozent-erklart>, zuletzt gesehen am 23.02.2013.

⁴⁰ z. B. Sprachkompetenzprojekt an der Universität Duisburg-Essen; Pilotprojekt „Lesen bildet (auch Juristen!)“ an der Universität Bochum; LiKom (Lese- und Schreibentwicklung von BA-Studierenden an der Universität Bielefeld etc.

DSH⁴¹, der TestDaF⁴² und/oder die Sprachprüfungen ZMP oder ZOP des Goethe-Instituts⁴³.

Die *DSH-Prüfung* ist eine sprachliche Zulassungsprüfung, die dazu dient, die für ein Hochschulstudium in Deutschland erforderlichen deutschen Sprachkenntnisse nicht muttersprachlicher Studienbewerber nachzuweisen. Die DSH-Prüfung wurde zunächst als Prüfung zum Nachweis deutscher Sprachkenntnisse (PNDS) bezeichnet. Sie ist auch von Studienbewerbern deutscher Staatsangehörigkeit abzulegen, die keinen deutschsprachigen Bildungsabschluss vorweisen können (z. B. Aussiedler mit deutscher Staatsangehörigkeit, doch ausländischer Hochschulzugangsberechtigung).

Die DSH-Prüfung umfasst zwei große Testteile: einen schriftlichen und einen mündlichen Teil. Mit dem schriftlichen Teil soll gewährleistet werden, dass Probanden über ausreichende Sprachfähigkeiten verfügen, um zum Beispiel folgende Studienvoraussetzungen zu erfüllen:

- Vorlesungen hörend zu verstehen, den Inhalt in Notizen festzuhalten und weiterzuverarbeiten,
- studienbezogene schriftliche Texte zu verstehen und sich damit auseinanderzusetzen,
- sich selbstständig schriftlich und mündlich zu einem vorgegebenen studienbezogenen Thema zu äußern⁴⁴.

Die schriftliche Teilprüfung umfasst folgende Aufgabenbereiche:

- Verstehen und Verarbeiten eines Hörtextes (*Hören*)
- Verstehen und Bearbeiten eines Lesetextes (*Lesen*)
- vorgabenorientierte Textproduktion (*Schreiben*)
- Verstehen und Bearbeiten wissenschaftssprachlicher Strukturen (*Verstehen*)⁴⁵

⁴¹ DSH = Deutsche Sprachprüfung für den Hochschulzugang.

⁴² TestDaF = Test Deutsch als Fremdsprache.

⁴³ ZMP oder ZOP = Zentrale Mittelstufen Prüfung oder Zentrale Oberstufen Prüfung des Goethe-Instituts.

⁴⁴ http://www.fadaf.de/de/sprachpr_fungen/dsh/, zuletzt gesehen am 27.05.2013.

⁴⁵ Zum Bestehen der DSH-Prüfung müssen beide Prüfungsteile (der schriftliche und der mündliche Teil) bestanden werden. Zum Erreichen der Niveaustufe DSH 2, die die Prüfungsteilnehmenden bereits zur Annahme des Studiums an einer Hochschule berechtigt, müssen beide Prüfungsteile mindestens mit dem Ergebnis DSH 2 bestanden werden. Die Bestehensgrenze für jede Teilprüfung liegt bei 57 %. Wird das Ergebnis zwischen 67 % und 81 % erreicht, wird die Niveaustufe DSH 2 und ab 82 % das Niveau DSH 3 bescheinigt.

Diese Aufgabenbereiche können unter Umständen auch kombiniert geprüft werden. Im Bereich Lesen etwa wird untersucht, ob wesentliche Inhalte eines vorgelegten Textes in ihrer spezifischen wissenschaftssprachlichen Struktur verstanden und verarbeitet werden können. Bei einem Lesetext handelt es sich um einen weitgehend authentischen Text, der die für wissenschaftliche Texte charakteristischen Strukturen aufweist und sich inhaltlich an wissenschaftlichen Fragestellungen orientiert, ohne jedoch spezielle Fachkenntnisse der Prüfungskandidaten vorauszusetzen.

Bewertet werden inhaltliche und sprachliche Aspekte in Abhängigkeit von der jeweiligen Aufgabenstellung, wobei insgesamt die Bewertung des Inhalts überwiegt. Außerdem fließt die Vollständigkeit und Angemessenheit der Erfüllung der gestellten Aufgaben in die Bewertung ein.

In der mündlichen Prüfung wird geprüft, ob Prüfungskandidaten allgemein verständliche Sachverhalte aus dem wissenschaftlichen Bereich erfassen und referieren bzw. kommentieren können. Außerdem wird überprüft, ob sie in der Lage sind, in einem Gespräch über studienrelevante Sachverhalte angemessen zu reagieren. Als Prüfungsmaterialien können in der Prüfung entsprechende Texte, Grafiken oder Schaubilder verwendet werden. Bewertet werden die allgemeine Gesprächsfähigkeit (Aufgaben- oder Fragenverständnis, angemessenes Reagieren) sowie die Fähigkeit, Sachverhalte korrekt und verständlich darzustellen⁴⁶.

Eine weitere Prüfung für ausländische Studienbewerber, deren Bestehen zur Aufnahme eines Studiums an einer deutschen Hochschule berechtigt, ist die *TestDaF-Prüfung*. Sie ist für diejenigen gedacht, die ihre Studienqualifikation nicht an einer deutschsprachigen Einrichtung erworben haben und vor Beginn des Studiums an Hochschulen in Deutschland entsprechend den Regelungen in den Hochschulgesetzen der Länder hinreichende deutsche Sprachkenntnisse nachweisen müssen. In der TestDaF-Prüfung werden vier Fertigkeiten geprüft: Leseverstehen, Hörverstehen, schriftlicher und mündlicher Ausdruck. Die Prüfungsergebnisse werden einer von drei Niveaustufen zugeordnet:

⁴⁶ Ein Beispiel der mündlichen DSH-Prüfung kann unter folgender Adresse angeschaut werden: <http://www.uni-potsdam.de/zessko-sprachen/daf/pruefung/dsh-muster/dsh-sprechen.html>, zuletzt gesehen am 27.05.2013.

- TestDaF-Niveaustufe 5 (TDN 5)
- TestDaF-Niveaustufe 4 (TDN 4)
- TestDaF-Niveaustufe 3 (TDN 3)

Die Beschreibung der sprachlichen Fähigkeiten zum Beispiel für den Bereich *Leseverstehen* sieht folgendermaßen aus⁴⁷: Der Prüfungskandidat

TDN 5: Kann geschriebene Texte aus dem studienbezogenen Alltag sowie Texte zu fächerübergreifenden wissenschaftlichen Themen, die sprachlich und inhaltlich komplex strukturiert sind, in ihrem Gesamtzusammenhang und ihren Einzelheiten verstehen und diesen Texten auch implizite Informationen entnehmen.

TDN 4: Kann geschriebene Texte aus dem studienbezogenen Alltag sowie Texte zu fächerübergreifenden wissenschaftlichen Themen, deren Struktur sich an der Allgemeinsprache orientiert, in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen.

TDN 3: Kann geschriebene Texte aus dem studienbezogenen Alltag in ihrem Gesamtzusammenhang und in wesentlichen Einzelheiten verstehen; kann Texte zu fächerübergreifenden wissenschaftlichen Themen in Teilen verstehen.

Im Bereich Lesen soll der Prüfungskandidat zum Beispiel zeigen, dass er geschriebene Texte, die im hochschulbezogenen Kontext relevant sind, angemessen rezipieren kann. Vorgelegt werden drei Texte unterschiedlichen Schwierigkeitsgrades, die keine Fachkenntnisse voraussetzen (vgl. zum Bereich Lesetests Kap. 5.4). Im Bereich schriftlicher Ausdruck wird geprüft, ob der Kandidat Schreibhandlungen, die im hochschulbezogenen Kontext relevant sind, angemessen ausführen kann (vgl. TestDaF-Prüfungsordnung, S. 6). Die schriftliche Leistung wird anhand eines geschriebenen Textes zu einem bestimmten Thema bewertet, zu dem der Kandidat eine Grafik mit Informationen zum Thema und Aussagen bzw. Fragen zu seinen Aspekten erhält. Die Aufgabenstellung verlangt die Beschreibung der Grafik und eine begründete Stellungnahme der zur Diskussion gestellten Aussagen bzw. Fragen. Die Bewertung des schriftlichen Ausdrucks erfolgt nach den Kriterien *Gesamteindruck, Behandlung der Aufgabe* und *sprachliche Realisierung*.

Die Stufen TDN 4 und TDN 5 reichen für die Aufnahme eines Hochschulstudiums in Deutschland. Die TestDaF-Prüfung gibt keine Auskunft über Stärken und

⁴⁷ Die Beschreibung weiterer sprachlichen Bereiche kann auf der Webseite des TestDaF-Instituts nachgelesen werden: http://www.testdaf.de/teilnehmer/tn-info_nivea.php, zuletzt gesehen am 28.05.2013.

Schwächen in verschiedenen Komponenten der Sprachkenntnis einer Testperson und liefert auch keine Hinweise auf Sprachbereiche, die gezielt zu fördern oder in ihrer Entwicklung zu beobachten wären (vgl. Eckes 2010a, S. 126).

Unter Prüfungen mit dem gleichen Ziel wären Sprachprüfungen des Goethe-Instituts (ZMP und ZOP)⁴⁸ zu nennen. Auch sie ermöglichen den Hochschulzugang in Deutschland und setzen ein hohes, sehr weit fortgeschrittenes Niveau voraus (C1-C2 nach GeR). Den genannten Tests ist die differenzierte Erhebung der Sprachteilkompetenzen gemein⁴⁹.

Das Bestehen der genannten Prüfungen setzt voraus, dass die Testperson über ein differenziertes Sprachwissen verfügt und auf unterschiedliche Komponenten dieses Wissens zugreifen kann (orthografische, lexikalische, morphologische, syntaktische, semantische und kontextuelle Aspekte). Die genannten Wissenskomponenten sind basale Sprachfähigkeiten, über die sowohl inländische Schulabsolventen als auch ausländische Studienbewerber verfügen sollten.

Die dargestellten Prüfungsformate für die ausländischen Studienbewerber erfassen einzelne Dimensionen der Sprachkompetenz. Aufgaben zu den Bereichen *Hören* und *Sprechen* wären im Rahmen der vorliegenden Studie für den universitären Bereich aus ökonomischen Gründen nicht realisierbar – vor allem, wenn große Gruppen getestet werden sollen. Für die Überprüfung der Schreibkompetenz bräuchte man zum einen andere Schreibimpulse, die an das muttersprachliche Niveau der Studierenden angepasst sind und zum anderen Bewertungsraster, die auf verschiedene Texte übertragbar sind, um den Studierenden schnelles Feedback zu geben. Um objektive Ergebnisse zu erzielen, würden Codierer benötigt, die für die Bewertung der Schreibaufgabe geschult werden müssten. Das Testformat für die Überprüfung der Lesefähigkeiten von Studierenden kann von der DSH- oder TestDaF-Prüfung übernommen werden. In beiden Prüfungen werden zur Überprüfung des Leseverständnisses Multiple-

⁴⁸ Dazu ausführlich in Kap. 5.3.

⁴⁹ Zum Vergleich beider Prüfungen (DSH und TestDaF) bei Krekeler, 2002, http://www.uni-due.de/imperia/md/content/elise/ausgabe_2_2002_krekeler.pdf, zuletzt gesehen am 27.05.2013.

Choice-Tests eingesetzt. Ihr Einsatz hat sich in der Praxis in der Durchführung und Auswertung als ökonomisch erwiesen. Die Grundlagentexte sollten dem Niveau der Studierenden angepasst werden. Allgemeine Texte eignen sich vermutlich nicht, da die Studierenden Fragen durch Vorwissen lösen könnten. Falls populärwissenschaftliche Texte zum Einsatz kommen, sollten Multiple-Choice-Tests mit erweiterten Antwortoptionen genutzt werden, da die Ratewahrscheinlichkeit durch das Hinzufügen der Antwortkategorie (*der Text sagt nichts dazu*) erschwert wird (vgl. dazu auch Kap. 5.4.1).

Zusammenfassend lässt sich festhalten, dass die vorhandenen Testformate bzw. Prüfungsaufgaben entweder nicht altersadäquat sind oder eine andere Zielsetzung haben. Ohne vorherige Modifikation der Aufgaben bzw. Tests scheint der Einsatz bei der Gruppe der Studierenden aus zwei Gründen problematisch: Verwendet man die Tests oder Aufgaben, die als geeignet einzuschätzen sind (z. B. aus dem Fremdsprachenbereich), so erzielt man nicht genügend Varianz in der Gruppe, das heißt, die Tests sind in der Regel zu leicht: Mehr als die Hälfte der zu testenden Probanden lösen die Aufgabe zu über 90 %. Außerdem gibt es keine Vergleichswerte bezüglich der Überprüfung der Gütekriterien in der Gruppe der Studierenden. Funktioniert der Test bei einer jüngeren Gruppe der Probanden und hat er ausreichende Reliabilitätswerte (Zuverlässigkeit in Bezug auf die Messung), so bedeutet das nicht, dass beim Einsatz bei einer höheren Altersgruppe ähnliche Werte erzielt werden. Die Reliabilität kann sinken, was wiederum Auswirkungen auf die Gültigkeit (Validität) des Tests und der Testergebnisse haben kann. Daher sollten für die Studierenden Tests entwickelt werden, die für diese Gruppe geeignet sind. Falls Tests oder Aufgaben aus anderen Studien übernommen werden, sollten diese im Vorfeld vorgetestet werden, um die Gütekriterien zu überprüfen.

4. C-Tests

Da C-Tests das zentrale Testinstrument der hier dargestellten Untersuchung sind, wird dieses Format nachfolgend stärker in den Fokus genommen. Bevor die Hauptidee dieses Testverfahrens erläutert wird, soll zunächst auf das Thema *Redundanz* eingegangen werden, weil dieser Aspekt für das Verständnis des C-Testprinzips von besonderer Bedeutung ist.

4.1 Kommunikationstheorie und Redundanz

Kommunikation findet täglich und überall auf der Welt in unterschiedlichsten Formen statt. Dabei gibt es einen Austausch von Informationen zwischen Menschen und eine wechselseitige Rekonstruktion von Bedeutungen. Voraussetzung für gelungene Kommunikation ist die Fähigkeit auf beiden Seiten, die betreffende Information sinnvoll einzuordnen und zu interpretieren, d. h. sowohl vom Sender als auch vom Empfänger wird ihr ein Sinn bzw. eine Bedeutung gegeben.

Die *Kommunikationstheorie*, auch *Informationstheorie* genannt (Shannon & Weaver, 1949), wurde unter Führung der Bell Telephone Company entwickelt und beschreibt, inwieweit eine gesendete Nachricht einen Empfänger erreicht. Zwei wichtige Fragen, mit denen die Kommunikationstheorie sich beschäftigt, sind, wie die übertragenden Informationen mathematisch zu definieren bzw. zu messen sind und wie Fehler bei der Informationsübertragung korrigiert und Störquellen ausgeglichen werden können. Eine der wichtigsten Voraussetzungen für eine erfolgreiche Kommunikation ist aus technischer Sicht die störfreie Übertragung der zu vermittelnden Botschaft. Kommunikation wird hier als „das Entgegennehmen einer Nachricht von einem Sender, der den gleichen Zeichensatz zur Informationsübertragung benutzt, wie der Empfänger“ verstanden (Shannon, 1948). Information, die in diesem Prozess übermittelt wird, wird als „eine mathematisch berechenbare Größe zur Beschreibung von Signal- und Übertragungseigenschaften in technischen Systemen“ betrachtet (Shannon, 1948).

Die Kommunikationstheorie geht davon aus, dass jede beliebige Nachricht aus Information besteht, die in verschiedene Symbole (Nummern, Buchstaben, Piepstöne, Morsecode) übersetzt werden. Die Nachricht ist nur dann komplett und in diesem Sinne ‚ideal‘, wenn jedes Symbol die maximale Menge an Information übermitteln kann. Jede potenzielle Einmischung in die Kommunikation wird als ‚noise‘ oder Störung bezeichnet, die der Kommunikation schaden kann, z. B. ein angeborener Aussprachefehler des Sprechenden. Störungen sind schwer vorherzusagen; sie können jederzeit und an jeder beliebigen Stelle des Kommunikationsprozesses auftreten. Die Nachricht kann beim Empfänger in einer solchen Form ankommen, dass die Rekonstruktion des Sinnes unmöglich oder sehr aufwändig ist. Daher haben Ingenieure bestimmte Schutzmaßnahmen erdacht, die es erlauben, ein System so zu schützen, dass die Decodierung der Nachricht trotz Störungen noch immer möglich ist. In das System werden dazu mehr Symbole eingebaut als notwendig, so dass der Einfluss der Störfaktoren verringert wird. Diese zusätzlichen Informationen oder Symbole werden als *Redundanz* bezeichnet.

Der Begriff Redundanz lässt sich auch aus der Informationstheorie ableiten und bezeichnet dort, wie viel Information im Durchschnitt pro Zeichen in einer Informationsquelle mehrfach vorhanden ist. Eine Informationseinheit ist dann redundant, wenn sie ohne Informationsverlust weggelassen werden kann. Eine Kommunikation lässt sich durch redundante Informationen fehlertolerant über einen Informationskanal durchführen, da unter Umständen verlorengegangene oder verfälschte Teilinformationen vom Empfänger aus ihrem Kontext rekonstruiert werden können.

Nach Aussage der Kommunikationstheorie sind Redundanzen in der Sprache erstrebenswert, da der Kommunikationsprozess zu beliebiger Zeit an beliebiger Stelle gestört werden kann. Die mathematische Aufgabe von Shannon und Weaver war es, eine Balance zwischen Redundanz und Störungen in der Informationsübermittlung zu finden und zu minimieren, „as much redundancy as necessary but as little as possible“ (Klein-Braley, 1985, S. 2).

Spolsky übertrug die Kommunikationstheorie auf die Testtheorie und bewies, dass die Sprache ein sehr redundantes System ist. Zum Beispiel ist das Englische

nach Schätzungen von Shannon auf der Buchstabenebene zu 75 % redundant. In einer realen Kommunikationssituation begegnen wir Redundanzen in unterschiedlicher Form: schlechte Fotokopien, unleserliche Schrift, akustisch gestörte Telefonate, Sprachdefekte usw. Wir müssen sie minimieren, um den Kommunikationsprozess nicht zu stören.

4.2 Übertragung der Redundanz auf die Testtheorie

Überträgt man die Kommunikationstheorie auf die Testtheorie, dann hat man es mit zwei spezifischen Gruppen zu tun, deren sprachliche Fähigkeiten getestet werden: Muttersprachler und Fremdsprachenlerner. Der Unterschied zwischen ihnen wird deutlich, sobald Störungen in eine Nachricht bzw. einen Text eingebaut werden, zum Beispiel leises Vorlesen, dialektale Besonderheiten des Vorlesers oder schlechte Aufnahmequalität (vgl. Spolsky et al, 1968): Der Fremdsprachenlerner macht mehr Fehler dabei, Redundanzen zur Rekonstruktion der gestörten Nachricht auszunutzen. So folgt z. B. in der deutschen Sprache nach der Buchstabenkombination *sc* immer ein *h*. Die Information *es folgt ein h* wäre für einen Muttersprachler oder fortgeschrittenen Fremdsprachenlerner redundant, da ihm dies bekannt ist. Es gibt aber auch inhaltliche Redundanzen: In dem Satz *Sie erhalten von uns zusätzlich ein kostenloses Geschenk* etwa ist das Wort *kostenlos* redundant. Oder in *Er träumte von einer großen Villa* wäre das Adjektiv *groß* redundant. Spolsky erklärt den Unterschied zwischen Muttersprachlern und Fremdsprachenlernern so, dass die muttersprachliche Fähigkeit, mit reduzierter Redundanz zu arbeiten, dem Fremdsprachenlerner auf unterschiedlichen Niveaus fehlt bzw. noch nicht so weit entwickelt ist, sei es in Bezug auf phonologisches, grammatisches, lexikalisches, semantisches oder anderes Wissen (vgl. Spolsky, 1973, S. 170).

Die Idee, die Redundanz der Sprache für die Sprachdiagnostik auszunutzen, bildet die Grundlage für eine Gruppe neuartiger Sprachtests, die globale oder integrative Tests genannt werden (vgl. Kap. 2). Wenn Muttersprachler Störungen verschiedener Natur in einer Nachricht unbewusst vervollständigen können (vorausgesetzt, sie beherrschen die Sprache in ihrem Wortschatz, ihrem

Regelwerk und ihrem kulturellen Hintergrund und erkennen die Redundanzen), brauchen Fremdsprachenlerner komplette sprachliche Redundanzen und sogar vielleicht etwas mehr an Informationen, um gestörte Elemente zu rekonstruieren – z. B. langsames Sprechen, bessere Aussprache, verstärkte Gestikulation und Mimik oder Wiederholung des Gesagten (vgl. Wockenfuß 2009, S. 92).

Die Unterschiede zwischen beiden Gruppen lassen darauf schließen, dass die Kompetenzen in der jeweiligen Sprache umso entwickelter sind, je besser die Sprecherinnen die reduzierte Redundanz für den kommunikativen Zweck kompensieren können. Um zu erkennen, welches sprachliche Element in einer Nachricht fehlt, ist Kenntnis über die sprachlichen Regeln notwendig. Um die Nachricht zu komplettieren, muss der Empfänger (Muttersprachler oder Fremdsprachenlerner) seine Kompetenzen aktivieren. Nach Chomsky (1965) und Spolsky (1973) erfordern Testverfahren, die auf der Basis der reduzierten Redundanz entwickelt werden, besondere Techniken, um die Probanden in eine Situation zu versetzen, in der sie ihre Kompetenz bzw. Performanz vorführen können (vgl. dazu Kap. 1.3). Die Tests, die auf der Basis der Theorie der reduzierten Redundanz entwickelt werden, bauen bestimmte Störungen in eine Nachricht oder einen Text ein: grafische (visuelle) Störungen, akustische Störungen oder Störungen in Form von Lücken und Tilgungen⁵⁰.

4.3 Übertragung der Redundanz auf die Formate *Cloze-Test* und *C-Test*

4.3.1 Cloze-Tests

Die vielleicht bekannteste Anwendung der Theorie der reduzierten Redundanz in Tests ist der *Cloze-Test*. Er wird zum ersten Mal von Taylor (1953) erwähnt. Das Format lässt sich folgendermaßen beschreiben: In einem Textauszug wird jedes n-te Wort (die Anzahl der gelöschten Wörter variiert zwischen 5 und 10 Wörtern) komplett gelöscht und soll vom Probanden sinngemäß rekonstruiert werden. Als korrekt werden dabei die vom Autor benutzten Wörter verstanden oder (wenn

⁵⁰ Zur Gruppe der Tests zur reduzierten Redundanz gehören der C-Test, der Noise-Test, der Cloze-Test und das klassische Diktat.

die Testautoren dies vor dem Testen festlegen) alle Wörter, die dem Sinn und Inhalt des Textes nicht widersprechen. Der Unterschied zwischen dem Cloze-Test und anderen Fill-The-Gaps-Formaten besteht darin, dass der Cloze-Test nach einem bestimmten System funktioniert. So sind die Störungen unsystematisch eingebaut, so wie sie auch in einer Kommunikationssituation nach Shannon und Weaver vorkommen könnten.

Das Wort „cloze“ ist von „closure“ abgeleitet. Closure ist der zentrale Begriff der Gestalttheorie. Ihr Ansatz lässt sich wie folgt zusammenfassen: Unsere Wahrnehmung tendiert dazu, Gestalten zu komplettieren, d. h. stets von einem Ganzen auszugehen und die Summe einzelner Bestandteile zu addieren (vgl. Taylor 1953, S. 415).

Taylor's Pilotuntersuchungen (1953) zeigten, dass die Fähigkeit, systematische Lücken in einem Cloze-Test zu rekonstruieren, mit Fähigkeiten übereinstimmt, die der realen Kommunikationssituation gleichen können. Er fand in seinen Testergebnissen signifikante Unterschiede zwischen Muttersprachlern und Nicht-Muttersprachlern und schlug vor, den Cloze-Test als Testformat für die Überprüfung der Lesekompetenz in der Muttersprache zu verwenden. Weitere Untersuchungen wurden von Carroll et al. (1959), Darnell (1968) und Oller (1971) mit je unterschiedlichen Zielen durchgeführt. Oller untersuchte die Fragen der Schwierigkeitslevels, Scoring-Techniken usw. 1973 wandte er sich dem Thema Messgenauigkeit der Cloze-Tests zu. Auf der Basis unvollständiger Informationen musste der Proband die Lücke rekonstruieren und so den Sinn des Satzes wiedergeben. In diesem Prozess werden sowohl produktive als auch rezeptive Fertigkeiten bzw. Fähigkeiten in der Sprache aktiviert. Bei dieser Leseprobe formuliert der Proband Hypothesen, hat Erwartungen an den Text bzw. an eine konkrete Stelle im Text oder an die Informationen, die weiter im Text vorkommen. Im weiteren Verlauf verwirft er seine Hypothesen oder belegt sie mit Fakten aus dem Text. Wenn die Hypothese falsifiziert wird, muss eine neue gebildet werden, die wiederum mithilfe des Textes überprüft wird. Dieser Prozess gleicht offensichtlich dem Prozess des wissenschaftlichen Arbeitens und aktiviert mehrere Prozessebenen, die für den Tester in diesem Moment nicht sichtbar sind. Oller vertritt die Auffassung, dass diese vorformulierten und noch

von den Probanden zu überprüfenden Hypothesen „manifestation [...] of the [...] underlying competence“ seien (Oller, 1973, S. 114).

Empirisch gesehen zeigen Cloze-Tests sehr gute Werte bzgl. der Reliabilität und auch der Validität; gewöhnlich überprüft man die Validität mithilfe der Korrelationen zwischen den Cloze-Tests und anderen Testverfahren, z. B. dem Diktat (vgl. dazu Oller, 1979).

In den achtziger Jahren wurde das Testformat Cloze-Test kritisiert. Im Folgenden werden die Hauptkritikpunkte zusammengefasst, die von Klein-Braley (1981) und Alderson (1978) formuliert wurden:

- Um gute Reliabilitätswerte zu erreichen, werden bis zu 50 Items im Cloze-Test benötigt. Dadurch wird der Text sehr lang.
- In Abhängigkeit vom Startpunkt der Cloze-Test-Lücken im Text können die Cloze-Tests unterschiedlich schwierig sein, was wiederum Auswirkungen auf die Reliabilität und Validität hat (vgl. dazu auch Studien von Klein-Braley, 1981).
- Die Auswertungsmethoden sind nicht optimal: Wird die Methode verwendet, in der nur Originallösungen akzeptiert werden, ist der Test sogar für Muttersprachler schwer. Verwendet man die Auswertungsmethode, bei der vom Original abweichende Lösungen akzeptiert werden, ist die Auswertung sehr zeitaufwändig (dazu auch Darnell, 1970).
- Die Erwartung, dass die für Nicht-Muttersprachler entwickelten Cloze-Tests von Muttersprachlern komplett korrekt rekonstruiert werden, erfüllte sich nicht.

Aufgrund dieser Kritik gegenüber den Cloze-Tests wurden Hauptkriterien formuliert, die ein Test erfüllen muss, um in der Praxis als akzeptabel zu gelten. Sie gelten bis heute als Regeln für den Nachfolger des Cloze-Tests, der einen ähnlichen Namen trägt und als *C-Test* bekannt ist. Bei der Konstruktion eines C-Tests bearbeitet man nicht ganze Wörter wie beim Cloze-Test, sondern nur Wortteile (vgl. Kap. 4.3.2, Kap. 4.4). Dieses Verfahren schlägt sich auch im Namen nieder, da von „Cloze“ lediglich der Anfangsbuchstabe („C“) geblieben ist.

Die Hauptkriterien lassen sich wie folgt formulieren (vgl. Raatz, 1985, S.14):

- 1) Der neue Test muss wesentlich kürzer sein und gleichzeitig 100 Items enthalten.

- 2) Die Lückenquote und der Startpunkt der Störelemente muss fixiert sein und immer an der gleichen Stelle beginnen, um persönliche Präferenzen zu vermeiden und den Test objektiv zu machen.
- 3) Die gelöschten Wörter müssen repräsentative Beispiele der Textelemente sein.
- 4) Probanden mit besonderen Kenntnissen sollen nicht durch spezifische Texte bevorzugt werden. Der neue Test muss verschiedene Texte enthalten.
- 5) Nur exakte Auswertung ist möglich, um Objektivität zu erreichen.
- 6) Muttersprachler sollen in der Lage sein, den Test zu 100 % rekonstruieren zu können.
- 7) Der Test muss reliabel und valide sein.

4.3.2 C-Tests

Wie bereits erwähnt, entstanden die C-Tests als Alternative zu den kritisierten Cloze-Tests, die u. a. wegen ihrer Tilgungstechnik und den daraus resultierenden Schwächen kritisiert wurden. Klein-Braley und Raatz (1985 a, b, c) arbeiteten an einem neuen Format und entschieden sich dafür, die Lücken nicht auf Satzebene, sondern Wortebene einzubauen. Der C-Test besteht aus mehreren kurzen Texten unterschiedlicher Thematik (in der Regel vier oder fünf)⁵¹, in denen in jedem zweiten Wort die zweite Hälfte des Wortes getilgt wird und sinngemäß rekonstruiert werden muss. Dieses Tilgungsprinzip wird ‚klassisch‘ oder auch ‚kanonisch‘ genannt und trägt in dieser Untersuchung zur besseren Orientierung den Namen *2er-Tilgung*, da auch noch andere Varianten von C-Test-Formaten vorgestellt und besprochen werden.

Der Hauptunterschied des C-Tests zum Cloze-Test liegt darin, dass den Probanden für die Bearbeitung mehrere Texte vorliegen und nicht nur einer. Sie müssen nur Wortteile rekonstruieren und nicht ganze Wörter. Im Folgenden wird auf einige Vorteile von C-Tests gegenüber den Cloze-Tests eingegangen.

- Das C-Test-Prinzip ermöglicht den Einbau mehrerer Items (Lücken) in kürzeren Texten. Der klassische Cloze-Test, in dem jedes fünfte Wort gelöscht wird, müsste 500 Wörter enthalten, damit der Test 100 Items enthält. Das würde einer DIN-A4-Seite pro Cloze-Test entsprechen. Der C-

⁵¹ Es ist möglich, mehr als fünf Texte in der Testbatterie zu verwenden; allerdings ist dabei zu beachten, dass das Testkonstrukt entsprechend länger ausfällt, was Auswirkungen auf die Konzentration der Testteilnehmer und die Ökonomie des Testverfahrens haben kann.

Test, der 20 Lücken enthält und aus 4 bzw. 5 kleinen Texten besteht, würde ebenso eine DIN-A4-Seite beanspruchen.

- Die C-Test-Auswertung ist präziser und objektiver, da nur eine bzw. wenige Lösungen akzeptiert werden. In manchen Fällen können zwar auch vom Original abweichende Lösungen akzeptiert werden. Allerdings sinkt die Anzahl der möglichen Varianten, da Wörter im Vergleich zum Cloze-Test nur zur Hälfte gelöscht werden.
- Die C-Test-Auswertung ist schnell und für muttersprachliche Auswerter einfach. Oft brauchen sie den Text nur einmal zu lesen und benötigen keinen Lösungsschlüssel.
- Für Muttersprachler ist die Rekonstruktion von C-Tests, die für den DaF-Bereich entwickelt werden, einfach. In vielen Pilotstudien wurde bestätigt, dass Muttersprachler solche C-Tests bis zu 95 % richtig lösen (vgl. Baur & Spettman, 2006, 2009).
- Bei der Tilgung jedes zweiten Wortes im Satz ist die Wahrscheinlichkeit der Repräsentativität verschiedener Wortklassen viel höher als beim Cloze-Test. Bei einer Überprüfung können durch Umstellungen oder durch das Einfügen einzelner Wörter ohne Aufwand Korrekturen vorgenommen werden.

4.4 Entwicklung eines klassischen C-Tests

Die Entwicklung und Auswertung eines klassischen C-Tests für eine vorgegebene Probandengruppe ist relativ einfach. Die gesamte Testbatterie, auch *Testset* oder *Testlet* genannt, soll aus vier Texten bestehen und 20 bzw. 25 Lücken pro *Teiltest* enthalten. Als *Teiltest* wird jeder einzelne C-Test in der Testbatterie bezeichnet. Die Durchführungszeit beträgt 20 Minuten, also 5 Minuten pro C-Test. Raatz und Klein-Braley (1985, S. 96ff.) haben Regeln für die Entwicklung des klassischen C-Tests formuliert, die ich im Folgenden gekürzt wiedergebe. Sie helfen, bei der Testentwicklung grobe Fehler zu vermeiden und sich an das klassische Format zu halten:

- a) Es werden 5 Texte von ca. 60-70 Wörtern benötigt. Der ausgewählte Textabschnitt soll in sich abgeschlossen sein. Die Grundlagentexte sollen inhaltsneutral sein, keinen speziellen Wortschatz enthalten oder spezielles Wissen erfordern. Für L1- und L2-Lerner, die mit ihrer Sprache täglich konfrontiert werden, können authentische Texte gewählt werden. Dasselbe betrifft fortgeschrittene Fremdsprachenlerner. Für weniger

fortgeschrittene Lerner sollen die Texte angepasst werden; dies erfolgt beispielsweise durch curriculaadäquate, adaptierte Texte oder einige 'Korrekturen' vom Lehrer, z. B. Ersetzung eines unbekannten Wortes durch ein Synonym etc. Es wäre unfair, die Lerner mit Texten zu konfrontieren, in denen Vokabeln oder grammatische/syntaktische Strukturen vorkommen, die ihnen unbekannt sind.

- b) Der erste Satz soll unbearbeitet bleiben. Die Tilgungen werden erst im zweiten Satz eingebaut. Beim zweiten Wort im zweiten Satz wird die Hälfte des Wortes getilgt, so dass die gelöschten Buchstaben durch einen langen Strich ersetzt werden. Wenn das Wort eine ungerade Anzahl von Buchstaben hat, wird mehr als ein Buchstabe gelöscht. Einbuchstabenwörter (z. B. Im Englischen *a* oder *I*), Namen, geografische Bezeichnungen, Zahlen und Abkürzungen werden bei Tilgungen übersprungen. Wenn die 20 Tilgungen eingebaut sind, wird der Satz ausgeschrieben, so dass der Text ein natürliches Ende hat.
- c) Die sechs Texte werden nach ihrer Schwierigkeit geordnet, diese kann hier noch nicht bestimmt werden. Die Einordnung erfolgt nach der Einschätzung der Lehrer oder Testentwickler.
- d) Für jede korrekte Lösung in einem C-Test wird ein Punkt vergeben. Pro Text können maximal 20 Punkte und im Gesamttest 100 Punkte erreicht werden.
- e) Der entwickelte C-Test soll etwa 10 erwachsenen Muttersprachlern vorgelegt werden. Erwartet wird, dass jeder einzelne von ihnen zu 95 % gelöst werden kann. Bei dieser Überprüfung kann es vorkommen, dass für einige Lücken weitere akzeptable Lösungen gefunden werden. Diese können als alternative Varianten für die korrekten Lösungen aufgenommen werden. Die C-Tests, in denen muttersprachliche Probanden weniger Punkte erreichen, sollen aus der Testbatterie entfernt werden⁵². Ein neuer C-Test kann als Ersatz erstellt werden (s. den Unterpunkt a).
- f) Nach der Auswertung wird der Test, wenn notwendig, an einer größeren Gruppe erprobt (z. B. 100 Probanden).
- g) Um zwei ungeeignete Tests zu identifizieren, werden Testschwierigkeiten (p-Werte) für jeden einzelnen Teilttest berechnet und untereinander mit dem Gesamtwert korreliert. Hier sind die Formeln dazu:

⁵² Das betrifft nur C-Tests, die für den DaF- oder DaZ-Bereich entwickelt werden.

(1)

$$p = \frac{x_j}{n_j} \cdot 100$$

(2)

$$r_{jt} = \frac{\sum i (x_{ji} - \bar{x}_j) \cdot (x_i - \bar{x})}{(N - 1) \cdot s_j \cdot s}$$

(3)

$$s_j = \sqrt{\frac{\sum i (x_{ji} - \bar{x}_j)^2}{N - 1}}$$

x_{ji} Anzahl der Punkte von der Person i im Text j

x_i Gesamtscore von der Person i

\bar{x}_j der durchschnittliche Testwert im Text j

\bar{x} der durchschnittliche Gesamtscore

s_j Standardabweichung im Text j

n_j Anzahl der Lücken (Tilgungen) im Text j

N Anzahl der getesteten Personen

(4)

$$r_{tt} = \frac{c}{c - 1} \cdot \left(1 - \frac{\sum s_j^2}{s^2} \right)$$

c - Anzahl der verwendeten Texte bzw. Teilttests

h) C-Tests, die unzureichende p-Werte aufweisen und als zu leichte oder zu schwierige C-Tests einzuschätzen sind, werden aus der C-Testbatterie entfernt. Als akzeptabel gelten in der Testpraxis p-Werte zwischen 0.30 und 0.70 (vgl. ALTE, Modul 2). Welche Schwierigkeitswerte noch als akzeptabel gelten können, hängt auch vom Zweck der Testung ab.

i) Danach wird die Reliabilität des entwickelten Testsets vor und nach dem Entfernen der "ungeeigneten" C-Tests überprüft. Der berechnete Cronbachs-Alpha-Koeffizient darf Werte zwischen $\alpha = 0.8$ und 0.9 annehmen, um für die Testzwecke als gut bzw. akzeptabel zu gelten.

j) Im Weiteren werden C-Tests nach ihrer Schwierigkeit geordnet, beginnend mit dem leichtesten C-Test.

- k) Der Gesamtscore ist bei der Verteilung der Probanden in verschiedene Lernergruppen oder zur Ermittlung eines Gesamteindrucks in Bezug auf eine Gruppe hilfreich. Normwerte und Korrelationskoeffizienten für verschiedene Probandengruppen können gemäß der klassischen Testtheorie berechnet werden (s. dazu das Kap. 7).

Im Folgenden wird auf weitere C-Test-Formen eingegangen, die in der Praxis Verwendung gefunden haben.

4.5 Modifizierte C-Tests

In der C-Test-Praxis wurden auch C-Test-Formate entwickelt, die von der dargestellten klassischen Form abweichen. Dazu zählen in erster Linie Abweichungen von der 2er-Tilgung und Abweichungen von der Endtilgung. Ich erläutere dies kurz am Beispiel der von Baur und Spettmann entwickelten *teilkfertigungsorientierten C-Tests*, der sogenannten TF-Tests: Bei ihnen wird die erste Worthälfte getilgt, sodass die Überprüfung der semantisch-lexikalischen Fähigkeiten fokussiert wird. Diese C-Tests werden oft in Verbindung mit fachbezogenen Texten eingesetzt (z. B. aus dem Fach Biologie, vgl. den C-Test *Das Gebiss des Wolfes* im Aufsatz von Baur & Spettmann, 2009, S. 117).

Die TF-Tests eignen sich zum Einführen und Einüben des Fachwortschatzes und zur Überprüfung des Leseverständnisses bei Fachtexten. Aber auch in Kombination mit verschiedenen grammatischen Themen, z. B. Einübung der (un)bestimmten Artikel, Personalpronomen oder Präpositionen, ist ein Einsatz von TF-Tests denkbar.

Es gibt einige Änderungen, die bei Worttilgungen in den TF-Tests beachtet werden müssen. Bei ungerader Buchstabenanzahl im Wort wird ein Buchstabe mehr (ggf. auch weniger) getilgt (z. B. Erwachsener oder Erwachsener). Da im Deutschen viele Komposita vorkommen, wird nur die erste Hälfte des ersten bedeutungstragenden Wortes getilgt (z. B. Fahrrad, Innenstadt oder Innenstadt). Beachtet man diese Regel nicht und tilgt die Hälfte des Wortes, ist die Rekonstruktion in manchen Fällen zu schwierig. Wie Buchstabenfolgen (sch, tsch, ck, ch) oder Diphthonge (ei, eu) getilgt werden, kann variieren und muss in einzelnen Fällen entschieden werden. Fällt die Tilgung auf die Buchstabenfolge

tsch und ist die Lücke aus dem Kontext nicht rekonstruierbar, kann man weniger Buchstaben löschen (zum Beispiel rutschen oder rutschen). Die Anzahl der getilgten Buchstaben hängt stark vom Ziel der Testung und von der getesteten Gruppe ab. Besteht die Untersuchungsgruppe aus jüngeren Probanden, können weniger Buchstaben getilgt werden als bei älteren Probanden (z. B. Studierenden). Die Untersuchungen zum Einsatz fachspezifischer TF-Tests in der Schule werden bei Baur & Spettmann, 2009 und im Berufskolleg bei Baur, Spettmann & Mashkovskaya (2010) beschrieben.

Die Tilgungsform, die den TF-Tests zugrunde liegt, wird auch *Wortanfangstilgung* oder *Stammtilgung* genannt. C-Tests mit Wortanfangstilgung sind auch unter dem Namen *X-Tests* bekannt (vgl. Sigott & Köberl, 1993). Erste Versuche, diese C-Tests an erwachsenen deutschen Muttersprachlern zu erforschen, gab es im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* an der Universität Duisburg-Essen (vgl. Scholten-Akoun & Baur, 2012; Scholten-Akoun, Kuhnen & Mashkovskaya, 2012; Bremerich-Vos et al., 2013). Möchte man die Wortschatzbreite, das Fachvokabular und gezielt das Leseverständnis messen, eignen sich dafür die C-Tests mit der Wortanfangstilgung. Da diese C-Tests bei der Rekonstruktion der Wörter primär die lexikalisch-semantischen Strategien fokussieren, differenzieren sie gut in der Gruppe der Studierenden. Die grammatisch-morphologischen Fähigkeiten, die sich durch den Einsatz von klassischen C-Tests abbilden lassen, spielen hier keine Rolle. Das ist der wichtigste Unterschied zu C-Tests mit Wortendtilgung. Daher eignen die klassischen C-Tests sich weniger zur Verwendung bei fortgeschrittenen Lernern oder bei erwachsenen Muttersprachlern.

Eine andere Modifikation von C-Tests betrifft die Bearbeitungszeit. Manipuliert man die Zeit, die den Probanden zur Verfügung steht, hat dies Auswirkungen auf die Schwierigkeit von C-Tests. So können zum Beispiel klassische C-Tests mit 2-er-Tilgung bei erwachsenen Muttersprachlern eingesetzt werden, für die eine Zeitlimitierung vorgesehen ist (vgl. dazu Untersuchungen von Wockenfuß, 2009; Grotjahn, 2009; Grotjahn, Schlak & Aguado, 2009). C-Tests mit Zeitlimitierung werden auch *Speed-C-Tests* (S-C-Tests) genannt. Sie messen vor allem die

Verarbeitungsgeschwindigkeit. Allerdings sind die S-C-Tests keine reinen Speed-Tests.

Bei klassischen Speed-Tests werden den Probanden sehr leichte Aufgaben vorgelegt, die sie ohne Zeitdruck prinzipiell lösen könnten. Die Zahl der gelösten Aufgaben gilt als Maß für die Schnelligkeit der Informationsverarbeitung auf der Grundlage eines Zeitlimits. Bei klassischen *Niveautests*, auch *Power-Tests* genannt, geht es um die Bearbeitung schwieriger Aufgaben ohne expliziten Zeitdruck. Daher handelt es sich bei C-Tests für erwachsene Muttersprachler um eindeutige Niveautests. Beim Einsatz von S-C-Tests werden den Probanden relativ schwierige Aufgaben vorgelegt. Außerdem wird die Leistung wie beim klassischen C-Test anhand der Anzahl der korrekt ausgefüllten Lücken gemessen. Es handelt sich beim S-C Test deshalb ebenfalls um einen Niveautest, jedoch mit einer mehr oder minder massiven Speed-Komponente. Die Anzahl der korrekt rekonstruierten Lücken stellt ein Maß dar sowohl für die Schnelligkeit der Informationsverarbeitung als auch für das Niveau der zur Bearbeitung notwendigen sprachlichen Kompetenzen (vgl. Grotjahn, Schlak & Aguado, 2009, S. 301f.).

Die Autoren gehen von der Hypothese aus, dass „die hochkompetente Verwendung von Sprache unter Echtzeitbedingungen sowohl eine breite deklarative (sprachliche) Wissensbasis als auch eine ausreichende Prozeduralisierung und Automatisierung des Wissens voraussetzt“ (Grotjahn, Schlak & Aguado, 2009, S. 302, vgl. dazu auch Untersuchungen von Grotjahn, 1997; Wockenfuß & Raatz, 2006). Wockenfuß & Raatz beschreiben in ihrer Studie zum Einsatz zeitlimitierter C-Tests bei erwachsenen Muttersprachlern, dass sich aufgrund theoretischer Überlegungen und empirischer Befunde im Hinblick auf schwierige muttersprachliche C-Test-Texte neben allgemeiner Sprachkompetenz und verbaler Intelligenz auch logisches Denken als spezielle weitere Leistungsdeterminante abbilden lässt (vgl. auch Raatz & Wockenfuß, 2006, S. 237; Wockenfuß, 2009).

In der vorliegenden Untersuchung werden C-Tests mit Wortanfangstilgung ohne Zeitlimitierung verwendet. Dafür gibt es zwei Gründe: Zum einen wird großer Wert darauf gelegt, dass die C-Tests in der Gruppe der Studierenden

augenscheinlich valide sind, d. h. Studierende zweifeln die Ernsthaftigkeit des C-Test-Formats und ihre Aussagekraft als Screening-Instrument in Bezug auf die Überprüfung der allgemeinen Sprachkompetenz nicht an. Zum anderen wird eine Differenzierung im unteren Kompetenzbereich angestrebt, da Befunde aus Studien zum Einsatz von C-Tests bei Studierenden belegen, dass vor allem in diesen Bereichen Studierende zu verorten sind, die gravierende sprachliche Defizite haben (vgl. Scholten-Akoun & Baur, 2012, Scholten-Akoun, Kuhnen & Mashkovskaya, 2012; Bremerich-Vos et al., 2013). Die S-C-Tests bieten die Möglichkeit, im oberen Kompetenzbereich stärker zu differenzieren; daher eignen sie sich nicht für die Verwendung in der vorliegenden Studie.

4.6 Einsatzmöglichkeiten von C-Tests und Gründe für den Einsatz im akademischen Bereich

Seit der ersten Veröffentlichung zum C-Test als innovatives und vielversprechendes Testformat (Klein-Braley & Raatz 1982; Klein-Braley 1984, 1985, Raatz 1985) hat sich in der C-Test-Praxis einiges geändert. Ich gehe nachfolgend auf einige Punkte ein, die für den heutigen Einsatz von C-Tests von Bedeutung sind und wichtige Veränderungen in der C-Test-Verwendung hervorheben: In den ersten Veröffentlichungen wurde behauptet, dass mithilfe der C-Tests reliable und valide Aussagen über die allgemeine Sprachkompetenz sowohl in der Muttersprache als auch in der Fremdsprache möglich sind (vgl. Raatz 1985, S. 18).

Die heutige Forschung im fremdsprachlichen Bereich belegt, dass die C-Tests zuverlässig messen, da sie sehr hohe Reliabilitätswerte aufweisen (vgl. Kap. 4.8 zu den Gütekriterien von C-Tests) und dass sich die Validität, die für die Interpretation der Testergebnisse von zentraler Bedeutung ist, durch Außenkriterien überprüfen und absichern lässt (z. B. Lehrerurteile, Noten in sprachenrelevanten Fächern wie Deutsch oder Fremdsprache). Allerdings wurde die Validität von C-Tests im muttersprachlichen Bereich nicht für alle Altersstufen untersucht und gesichert (vgl. Wockenfuß, 2009). Die Verwendung von C-Tests

im Deutschen wurde weitgehend an monolingualen Schülern der 4.-8. Klassen untersucht (vgl. Publikationen von Baur & Spettmann, 2006-2012). Inzwischen wurden im Deutschen C-Tests nicht nur für Muttersprachler und Fremdsprachenlerner entwickelt, sondern auch für Zweitsprachsprecher (vgl. dazu Publikationen von Baur & Spettmann).

Ein weiterer Punkt betrifft den Umstand, dass die Testergebnisse in Bezug auf die sprachlichen Fähigkeiten der Lerner nur einen Gesamteindruck vermitteln und keine direkten Schlussfolgerungen bzgl. der Kompetenzentwicklung auf individueller Ebene erlauben. „They (c-tests) tell us nothing about specific features of grammar, or syntax, or vocabulary, or spelling, and they offer no direct evidence about performance in listening and speaking“ (vgl. Raatz, 1985, S. 18). Die Aussage, dass die klassischen C-Tests keine detaillierten Informationen über individuelle sprachliche Stärken oder Schwächen der getesteten Personen zulassen, trifft bei der Verwendung der C-Tests heutzutage nicht mehr zu – vor allem, wenn man die Verwendung der C-Tests in der Schule in den Blick nimmt, wie sie von Baur & Spettmann entwickelt wurden. Durch die von ihnen vorgeschlagenen C-Test-Auswertungen sind Aussagen zu Bereichen der rezeptiven und produktiven Kompetenz der Probanden möglich (vgl. Baur & Spettmann, 2006b, 2007).

Im schulischen Bereich, z. B. im Übergang von der Grundschule zur Sekundarstufe I, wurde mithilfe einer angepassten Auswertung in den Bereichen Sprachrichtigkeit und Worterkennung eine zweischrittige Analyse vorgenommen. Hier werden zum einen die Schüler und Schülerinnen in ein Gruppenranking eingeordnet, zum anderen wird ein individuelles Förderprofil erstellt. Im Zusammenhang mit der vorliegenden Arbeit ist von besonderem Interesse, dass durch die Analyse und Bewertung der Worterkennung die *Lesefähigkeit* der Schüler geprüft werden kann.

Baur, Grotjahn & Spettmann (2006b) haben eine Auswertungsmethode entwickelt, die es ermöglicht, nicht nur einen Gesamtüberblick über die Gruppe zu gewinnen, sondern auch Testergebnisse auf individueller Ebene zu interpretieren und zu differenzieren. Die Auswertung erfolgt nach zwei Kriterien:

Bewertet wird 1) der Bereich der *Worterkennung* (WE-Wert) und 2) der Bereich der *Sprachrichtigkeit* (R/F-Wert).

Im ersten Schritt wird analysiert, ob der Proband das Wort erkannt hat und im zweiten, ob es sprachlich richtig umgesetzt wurde. Die Differenz zwischen beiden Werten, die am Ende der Auswertung steht, auch Differenzwert (DF-Wert) genannt, spiegelt das Verhältnis zwischen den produktiven und rezeptiven sprachlichen Fähigkeiten eines Schülers wider. Dieser Wert fließt in die Bewertung der Testergebnisse, aber auch in die individuelle Beratung bezüglich einer sprachlichen Förderung ein, falls diese in Frage kommt (dazu Baur & Spettmann, 2006a, 2006b, 2007).

Dass die C-Tests keine Details zu Bereichen wie Grammatik, Syntax, Lexik und zu den Bereichen Hören und Sprechen liefern, wird teilweise auch heute noch bestätigt. Um Grammatik oder Syntax zu überprüfen, müssen andere Testverfahren verwendet werden. Da die C-Test-Texte schriftlich rekonstruiert und weder vorgelesen noch besprochen werden müssen, werden die Bereiche Hören und Sprechen nicht erfasst. Allerdings liegen auch Befunde vor, die den Schluss zulassen, dass es zwischen den C-Testergebnissen und den Ergebnissen aus Tests zum Hörverstehen und Sprechen Korrelationen gibt (vgl. Eckes & Grotjahn, 2006; Grotjahn, 2012). Diese Zusammenhänge werden so interpretiert, dass Sprache ein mehrdimensionales, komplexes Konstrukt ist und einzelne sprachliche Bereiche sehr eng untereinander verbunden sind. Im Folgenden wird auf Vorteile und Nachteile von C-Tests eingegangen.

Tabelle 4: Überblick über Vorteile und Nachteile von C-Tests

Vorteile	Nachteile
Durchführungsökonomie	Akzeptanz des Testformats (Augenscheinvalidität)
Auswertungsökonomie	globale Einschätzung der sprachlichen Fähigkeiten
Durchführungsobjektivität	Kaum weiter diagnostisch verwertbare Informationen ⁵³
Auswertungsobjektivität	Weitere Tests notwendig, um ausführliche Informationen zu den Teilkompetenzen zu bekommen
Kostenökonomie	
Fairness den Probanden gegenüber	Beschränkungen auf schriftsprachliche Teilfertigkeiten
Zumutbarkeit der C-Tests	
Transparenz der Instruktionen	
Vergleichbarkeit der C-Tests mit gleichen Schwierigkeiten (Parallelförmigkeit)	
Normierung und Skalierung ist für verschiedene Gruppe möglich	
Äußere Gestaltung ist ökonomisch und ansprechend	
Mögliche Modifikation des Testformats für verschiedene Zielgruppen	
Universalität/Flexibilität des Formats	

Ein großer Vorteil von C-Tests, der ihren Einsatz in der Praxis attraktiv macht, liegt in der Universalität und Flexibilität des Formats (vgl. Tabelle 4). C-Tests können an verschiedenen Institutionen und für unterschiedliche Zwecke eingesetzt und an verschiedene Gruppen angepasst werden. Ihr

⁵³ Das betrifft die Verwendung von C-Tests in der Gruppe der erwachsenen Muttersprachler oder fortgeschrittenen Fremdsprachenlerner, die über gut ausgebildete Sprachfähigkeiten verfügen. Bei jungen Lernern oder weniger fortgeschrittenen Fremdsprachenlerner ist eine zweistufige Auswertung von C-Testergebnissen möglich, die mehr diagnostische Informationen liefert (vgl. dazu Baur & Spettmann 2006-2012).

Verwendungsbereich hat sich in den letzten zehn Jahren sehr ausgeweitet, von der Screening-Funktion in Form eines Einstufungstests in der Fremdsprache bis hin zum C-Test als Förderinstrument, mit dem einzelne sprachlich besonders problematische Bereiche eingeübt werden können. Heutzutage werden C-Tests an verschiedenen Institutionen mit unterschiedlichen Zielen verwendet. Im Folgenden wird auf Einsatzbereiche und Ziele von C-Tests eingegangen.

1) C-Tests an Sprachschulen

Ziele: Einstufung, Selektion, Überprüfung des Lernfortschritts

C-Tests im Fremdsprachenbereich sind beliebt. Da sie primär auch für diesen Bereich entwickelt wurden, liegt eine große Anzahl von C-Tests für diese Testgruppe vor. Heutzutage stehen C-Tests in mehr als 25 verschiedenen Sprachen zur Verfügung, in europäischen und nichteuropäischen wie Chinesisch, Japanisch oder Russisch etc. Die C-Tests werden zur Überprüfung der allgemeinen Sprachkompetenz bzw. von sprachlichen Kenntnissen in einer Fremdsprache eingesetzt – etwa als Einstufungstests, um die Lerner verschiedenen Lerngruppen zuordnen zu können. Anhand der Testergebnisse mithilfe der Prototypengruppenmethode, die Eckes als Analyse vorgeschlagen hat, können C-Testergebnisse verschiedenen Kompetenzniveaus (GER) zugeordnet werden (dazu Eckes, 2010c). Mit den C-Tests lassen sich Fremdsprachenlerner testen und in Bezug auf das erreichte Sprachniveau verschiedenen Sprachgruppen zuordnen. Wird der C-Test an Eingangsvoraussetzungen geknüpft, kann er zum Beispiel zur Selektion von Studenten in einem bestimmten Studiengang verwendet werden. Setzt man den C-Test am Beginn und Ende eines Sprachkurses ein, kann mit ihm auch der Lernfortschritt gemessen werden – zum Beispiel im Bereich Lexik.

2) C-Tests an Schulen

Ziele: *Ranking, Diagnose, Förderung, Überprüfung des Lernfortschritts*

An Schulen lassen sich mit den C-Tests sprachliche Fähigkeiten in verschiedenen Klassenstufen erfassen. Erfolgreich entwickelt und eingesetzt wurden die C-Tests sowohl für niedrigere als auch für höhere Klassenstufen (vgl. Publikationen von Baur & Spettmann, 2006-2012). Da Schulklassen in der Regel gemischt sind, konnten C-Tests auch an bilingualen Schülern miterprobt werden. Die betreffenden Studien ermöglichten einen Vergleich der Testergebnisse zwischen zwei verschiedenen Lernergruppen, die in Abhängigkeit von der Muttersprache eigene Entwicklungsprozesse durchlaufen. Im schulischen Bereich konnten C-Tests auch als Lernstandserhebungsinstrument verwendet werden, d. h. mit ihrer Hilfe wurde der Lernstand dokumentiert, zum Beispiel am Anfang und Ende eines Schuljahres. Auf diese Weise konnten Fördergruppen gebildet werden, falls die Testergebnisse im Vergleich zum normierten Gruppendurchschnitt zu niedrig waren. Im schulischen Bereich wurde der C-Test zum ersten Mal als Diagnose- und Förderinstrument erprobt. Aufgrund der zweistufigen Bewertung (Sprachrichtigkeit und Worterkennung) konnte z. B. in den Klassenstufen vier bis sieben allgemeinsprachlicher Förderbedarf in bestimmten sprachlichen Teilbereichen (z. B. Verwechslung von langen und kurzen Vokalen, Fehler bei verdoppelten Konsonanten, Schwierigkeiten bei Groß- und Kleinschreibung) aufgedeckt werden. Als Fördermaterial kam der für die Klassenstufe geeignete vorerprobte C-Test in Frage, an dem Schüler festgestellte sprachliche Probleme übten. Von der erfolgreichen Verwendung von C-Tests als Förderinstrument berichten Baur & Spettmann (2007, 2008, 2009). Testet man sprachliche Fähigkeiten eines Fünftklässlers mithilfe geeigneter C-Tests am Anfang und Ende eines Schuljahres, kann der Lernfortschritt eingeschätzt werden.

3) C-Tests am Berufskolleg

Ziele: *Ranking, Diagnose, Förderung, Überprüfung des Lernfortschritts*

Seit kurzem gibt es auch erste Versuche zur Konstruktion und zum Einsatz von fachspezifischen C-Tests für Lerner am Berufskolleg (vgl. auch das Kap. 4.5). Diese Tests wurden mit dem Ziel entwickelt, Berufsschüler sowohl in fachsprachlichen als auch allgemeinsprachlichen Bereichen zu testen. Von den Autoren Baur, Mashkovskaya & Spettmann (2010) wurde die Hypothese überprüft, ob Berufsschüler Textverständnis sichernde Lesestrategien besitzen müssen, um die sich durch einen fachspezifischen Wortschatz und komplexere Strukturen auszeichnende Fachtexte semantisch korrekt erschließen zu können. Die verwendeten Tests wurden aus verschiedenen Texttypen kombiniert; sie bestanden aus vier C-Tests, entwickelt auf der Grundlage zweier allgemeinsprachlicher und zweier fachspezifischer Texte. In dieser Studie konnte beobachtet werden, dass Schüler mit einer guten allgemeinsprachlichen Lesefertigkeit auch bessere Leseleistungen erbringen. Schlechte fachsprachliche Ergebnisse ließen vermuten, dass diese auf Probleme mit dem Lernstoff zurückzuführen sind (z. B. unbekannter Wortschatz).

Durch die Ermittlung der WE-Werte wurden aussagekräftige Daten bezüglich des allgemeinen und fachspezifischen Textverständnisses und der Lesekompetenz der Schüler gewonnen. Außerdem erwies es sich als sinnvoll, das Verständnis von Fachtexten nicht abgekoppelt, sondern in Kombination mit dem allgemeinsprachlichen Textverständnis zu überprüfen und somit auch das Verhältnis zwischen beiden Bereichen zu untersuchen. Mithilfe der Differenzwerte zwischen R/F- und WE-Werten wurden wichtige Informationen für die sprachliche Förderung gesammelt (Baur, Mashkovskaya & Spettmann 2010, S. 26f., 35). In diesem Bereich wurden sowohl monolinguale als auch bilinguale Schüler getestet. Für diese Gruppe lassen sich TF-C-Tests (Teilfertigkeitstests) einsetzen (dazu ausführlich in Kap. 4.5).

4) C-Tests an Universitäten

Ziele: *Selektion, Ranking, Diagnose, Förderung*

Bis jetzt werden C-Tests an Universitäten vor allem im Bereich der Fremdsprachen als Einstufungstests für Studierende verwendet. Als Vorläufer der C-Tests für Muttersprachler des Deutschen können die von Baur & Spettmann zur Erfassung der Sprachkenntnisse von Zweitsprachenlernern entwickelten C-Tests angesehen werden. Erstmals wurden C-Tests für muttersprachliche Adressaten von Wockenfuß (2009) entwickelt (s. u.). Ihre Arbeit beeinflusste die Entwicklung der C-Tests für die Überprüfung der allgemeinen Sprachkompetenz für Lehramtsstudierende im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* an der Universität Duisburg-Essen. Hier wurden erstmalig C-Tests entwickelt, mit denen die Fähigkeiten erwachsener Muttersprachlern in größerem Umfang getestet wurden. Über die Entwicklung dieser C-Tests berichtet ausführlich Kap. 4.4. Mit solchen C-Tests lässt sich ein Ranking über die untersuchte Gruppe der Studierenden erstellen, und über normierte, standardisierte Werte ist ein Vergleich der Testergebnisse von Studierenden in verschiedenen Fachrichtungen oder an verschiedenen Universitäten möglich. Werden fachspezifische C-Tests z. B. am Anfang des Studiums eingesetzt, um eventuelle sprachliche Schwächen der Studierenden zu identifizieren, können C-Tests als Selektionsinstrument verwendet werden, falls z. B. im Anschluss daran eine Entscheidung über die Aufnahme eines Studierenden in einen Kurs getroffen wird. Kombiniert man C-Tests mit einer weiteren Aufgabe oder einem weiteren diagnostischen Test, lässt sich der so ermittelte Globalwert als Voreinschätzung der sprachlichen Leistungen im Sinne der prognostischen Validität (Vorhersagevalidität) verwenden (vgl. dazu Kap. 8.2).

Die ersten C-Tests an erwachsenen Muttersprachlern wurden, wie oben erwähnt, von Wockenfuß (2009) erprobt. Die C-Test-Präsentation wurde hier in besonderer Weise für Schüler modifiziert, indem sie zeitlich reduziert und an das sprachliche Niveau von Muttersprachlern angepasst wurde. Wockenfuß wollte

im Rahmen eines psychologischen Ansatzes überprüfen, ob die Lösungsgeschwindigkeit bei C-Test mit den kognitiven Fähigkeiten der Schüler korreliert, so dass die Tests auch in Kombination mit Intelligenztests angewendet werden können.

5) C-Tests in anderen Kontexten (Online-C-Tests)

Ziele: *Selektion, Ranking, Einstufung*

C-Tests können auch zu anderen Zwecken als den oben genannten eingesetzt werden, z. B. als Entscheidungshilfe für Weiterbildungsmaßnahmen im Bereich Sprache (Jakschik 1992, 1996), beim Fremdsprachenwettbewerb mit kompetenten Lernern (vgl. Grotjahn, Klein-Braley & Raatz 1992; Raatz & Klein-Braley, 1994) oder sogar mit Probanden mit Lernbehinderung als Teilaufgabe (z. B. KNUSPEL-L)⁵⁴. Das bestätigt noch einmal die Flexibilität und Universalität des Testformats.

Heutzutage setzt der C-Test sich als Computer-Version durch, was neue Nutzungsmöglichkeiten eröffnet⁵⁵. Beide Testvarianten (Papier-Version und Online-Version) wurden in Studien miteinander verglichen (Bisping & Raatz 2002; Sigott 2002); dabei wurden im Bereich Reliabilität und Validität gleich hohe Werte festgestellt. Die computergestützte Variante bietet eine umfangreiche Stichprobenbasis, die für die Normierung der Daten wichtig ist.

OnDaF verwendet seit einigen Jahren den deutschen C-Test in der Online-Variante, was das gleichzeitige Testen ausländischer Testteilnehmer aus verschiedenen Ländern möglich macht (vgl. dazu Eckes, 2010a, c). An mehreren Universitäten wird der C-Test in der Online-Form angeboten und dient hier als Einstufungstest für Sprachkurse oder für das Lehramtsstudium – zum Beispiel in den Fächern Englisch, Französisch, Spanisch etc.⁵⁶ In unserer Studie wird das

⁵⁴ Marx, H. (1998): Knuspels-Leseaufgaben (KNUSPEL-L): Handanweisung, Göttingen [u.a.]: Hogrefe.

⁵⁵ Im Rahmen des Sprachkompetenzprojektes „Diagnose und Förderung“ wurden C-Tests als Probe-Version an ca. 400 Studierenden erprobt.

⁵⁶ Hier wird eine Reihe der Universitäten aufgeführt, die den C-Test in der Online-Version anbieten: Augsburg, Berlin, Bielefeld, Bochum, Braunschweig, Chemnitz, Düsseldorf (IIK),

Papier-Format verwendet, doch inzwischen wurde auch eine Online-Version eingerichtet und im Rahmen großer Veranstaltungen an ca. 400 Probanden erprobt. In Bezug auf den Einsatz der Online-Version erscheint es uns wichtig, dass die Durchführung an der Universität in speziell dafür eingerichteten Räumlichkeiten⁵⁷ und unter Aufsicht durchgeführt wird, damit bei der Arbeit am Computer gegenseitiges Abschreiben vermieden wird⁵⁸ und kontrollierte Testbedingungen vorliegen. Dabei werden die Personen mit einem ID-Code identifiziert und bearbeiten die Texte nur so lange, wie die zeitlich voreingestellte Maske es zulässt⁵⁹. Im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* an der UDE wird angestrebt, mit mehreren Universitäten zu kooperieren und die Online-Version zu verwenden. Nicht jede Universität verfügt allerdings über mit Computern ausgestattete Räume, die es erlauben, 200-300 Studierende⁶⁰ gleichzeitig den Test bearbeiten zu lassen. Diese Einschränkung erschwert im Moment noch die vollständige Umstellung auf die Online-Version.

4.7 Kritik an C-Tests

Im Folgenden soll auf Kritik an C-Tests eingegangen werden, da der C-Test von Beginn seiner Einführung an sowohl von Theoretikern als auch von Praktikern

Erlangen, Duisburg-Essen, Gießen, Hohenheim, Kassel, Marburg, München, Münster, Paderborn, Passau, Rostock, Siegen, Stuttgart u.a.

⁵⁷ Die Universität Duisburg-Essen verfügt über spezielle Räumlichkeiten, die für Prüfungen mit Computern ausgestattet sind: <http://www.uni-due.de/zim/services/pc-pruefungen/pc-hall.shtml>, zuletzt gesehen am 13.06.2012.

⁵⁸ In der PC-Hall in Essen sind die Bildschirme so in die Tische eingebaut, dass das Abschreiben voneinander unmöglich ist.

⁵⁹ Bei der Verwendung der Papierversion ist eine Kontrolle der Bearbeitungszeit schwierig, besonders in großen Räumen: Beim Einsammeln der Testhefte kann es geschehen, dass einige Personen die Texte weiterbearbeiten, obwohl die Bearbeitungszeit zu Ende ist. Solche Fälle werden durch Einführung der Onlineversion vermieden, weil die Zeit hier durch den Computer kontrolliert wird. Wenn die Zeit abgelaufen ist, wird der Text automatisch ausgeblendet und steht nicht mehr zur Verfügung.

⁶⁰ Das kann der Größe eines einführenden Seminars im Grundstudium oder einer Einführungsveranstaltung zum Beispiel im Fach Pädagogik entsprechen; teilweise gibt es an der Universität Duisburg-Essen jedoch auch Veranstaltungen, die gleichzeitig von 1.000 Studierenden besucht werden. In einem solchen Fall ist die Online-Testdurchführung für alle Teilnehmer gleichzeitig nicht möglich. Es wird nach Möglichkeiten gesucht, wie eine Testung so großer Gruppen realisiert werden kann, ohne dass es zu Unregelmäßigkeiten kommt, welche die Testgütekriterien beeinträchtigen.

kritisiert wurde. Die dargestellten Aspekte fassen einzelne Kritikpunkte zusammen, die im Laufe der Jahre aufgetreten sind (und häufig auch ungeprüft wiederholt wurden) und gehen auf Kritikpunkte ein, die bei der Einführung des Screenings im Rahmen des Sprachkompetenzprojekts an der Universität Duisburg-Essen geäußert wurden⁶¹ (vgl. dazu auch die Publikation von Baur & Mashkovskaya (2013).

1. *Rückschlüsse auf die Sprachkompetenz einer Person lassen sich nicht aus C-Tests ableiten. Zur Messung der Sprachkompetenz benötigt man zeitaufwändige Testaufgaben zu einzelnen Teilkompetenzen.*

Der Anspruch, mit den C-Tests in kürzester Zeit globale Informationen über sprachliche Fertigkeiten der Testteilnehmer zu gewinnen, überzeugt viele Probanden und Testanwender nicht. Dieser Punkt wurde bereits 1994 von Freese hervorgehoben. Er betonte, dass der Begriff *Sprachkompetenz* nicht klar definiert ist und daher „die Interpretation von C-Test-Ergebnissen fast der Beliebigkeit anheimgegeben“ sei, „da nicht bekannt ist, was eigentlich damit gemessen wird“ (Freese 1994, S. 309). Dem kann entgegengehalten werden, dass gerade die Korrelationsuntersuchungen des C-Tests mit anderen evaluierten Tests, die einzelne Teilfertigkeiten messen, bestätigen, dass der C-Test die Sprachfertigkeiten in der Tat ‚global‘ erfasst. Mehrere Untersuchungen belegen zum Teil hohe Korrelationen mit verschiedenen Teilfertigkeiten, zumindest bei Fremdsprachenlernern (vgl. Klein-Braley, 1996; Eckes & Grotjahn, 2006a, b; Grotjahn, 2010a). Insofern wurde diese Kritik, die gern übernommen wird (vgl. Ehlich 2012), durch die empirische Forschung widerlegt. Entsprechende Untersuchungen für Zweitsprachenlerner und Muttersprachler stehen zwar noch aus; festzuhalten bleibt jedoch, dass diese ‚Sprachkompetenz‘-Kritik den Blick für die Erforschung des C-Tests geschärft hat.

⁶¹ Die gesammelten Kritikpunkte beruhen auf zahlreichen Aussagen von Studierenden der nordrhein-westfälischen Universitäten, deren sprachliche Fähigkeiten mit den C-Tests überprüft wurden und von einigen Dozenten, die die Testungen in ihren Seminaren im Rahmen der vorliegenden Studie und im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* an der Universität Duisburg-Essen ermöglicht haben.

Um Fremdsprachenlerner schnell und relativ zuverlässig verschiedenen Niveaugruppen zuordnen zu können, werden C-Tests zunehmend in dieser ‚globalen‘ Funktion eingesetzt. In der Praxis reicht hier ein kurzer Textausschnitt, um zu überprüfen, ob a) der Text verstanden wurde, b) die Rekonstruktion der Lücken orthografisch, grammatikalisch und semantisch korrekt erfolgte und c) eine Kategorisierung von Auffälligkeiten möglich ist. Die Theorie der reduzierten Redundanz postuliert, dass die Sprachkompetenz der Probanden umso höher ist, je vollständiger die Redundanzen in der Sprache bzw. in einem Text erkannt werden (vgl. dazu ausführlich Kap. 3.2). Die C-Test-Texte haben einen hohen Informationsgehalt und enthalten verschiedene sprachliche Phänomene, die auf der Grundlage sprachlichen Regelwissens erkannt und vervollständigt werden können. Um die Sprachkompetenz voreinzuschätzen, ist der Einsatz zuverlässiger und ökonomischer Testverfahren, zu denen der C-Test zu rechnen ist, durchaus zu empfehlen.

2. C-Tests können kein Leseverständnis messen, da sie durch die eingebauten Lücken keine rein-lineare Struktur haben.

Der Vorwurf, dass C-Tests nicht das Leseverständnis messen, ist eigentlich keine Kritik am C-Test, sondern eine Kritik an der Interpretation von C-Test-Ergebnissen als Hinweis auf die Lesekompetenz der Getesteten.

Wie ich in Kap. 6 ausführe, ist diese Kritik nur haltbar in Bezug auf die Testung von Fremdsprachenlernern und wenn als Testformat die Wortendtilgung gewählt wird. Die Überlegungen für Zweitsprachenlerner von Baur & Meder (1994) und Baur & Spettmann (2006b) weisen in eine andere Richtung und entkräften die im Kontext der Fremdsprachenforschung geäußerte Kritik (vgl. insbesondere Grotjahn und Tönshoff, 1992). Meine eigenen empirischen Untersuchungen mit Muttersprachlern zeigen, dass der C-Test für diese Adressatengruppe als Lesetest valide ist. Die linguistische und psycholinguistische Grundlage sei im Folgenden noch einmal kurz zusammengefasst.

Zur Rekonstruktion von C-Tests wird den Probanden ein kleiner, in sich abgeschlossener Textabschnitt vorgelegt. Um die Lücken sinngemäß zu vervollständigen, müssen auch die mit der Lesekompetenz verbundenen Fähigkeiten aktiviert werden. Da die Lücken der Wort-, Satz- und Textebene zugeordnet werden können (vgl. dazu ausführlich Kap. 5), erlauben niedrige bzw. höhere C-Testwerte Rückschlüsse auf verschiedene Prozessebenen der Lesekompetenz (vgl. auch Kap. 4 dazu). Die in der vorliegenden Studie eingesetzten C-Tests mit Wortanfangstilgung ermöglichen unter anderem auch die Überprüfung der Lesekompetenz, da sie im Vergleich zu den klassischen C-Tests stärker lexikalisch-semantische Fähigkeiten fokussieren. Rekonstruiert der Leser alle Lücken im C-Test, liegt ihm ein Text mit einer linearen Struktur vor. In der C-Test-Praxis ist bewiesen worden, dass C-Tests mit Tests zur Überprüfung der Lesekompetenz korrelieren (vgl. Kap. 5, Kap. 7). Ob sich diese Zusammenhänge auch bei muttersprachlichen C-Tests bestätigen lassen, wird im Rahmen dieser Studie überprüft (vgl. Kap. 7, Kap. 8).

3. Durch die globale Auswertung von C-Tests werden einige Probanden benachteiligt. Probanden, welche die Lücken z. B. semantisch falsch rekonstruieren, erzielen ähnliche Leistungen wie die, welche die Lücken unbearbeitet lassen. Daraus lassen sich keine Rückschlüsse auf die Sprachkompetenz der Probanden oder ihr Leseverständnis ableiten.

Zur Bewertung sprachlicher Leistungen in C-Tests können verschiedene Auswertungsmethoden angewendet werden. Die zweistufige Auswertungsmethode, die von Baur & Spettmann (2006) vorgeschlagen wird und bei der Bewertung von Sprachkompetenz zwei Werte liefert (WE- und R/F-Wert) (vgl. Kap. 4.6), kann bei Studierenden aus den nachfolgend genannten Gründen nicht verwendet werden: Zum einen basieren C-Tests auf der Wortanfangstilgung, die stärker lexikalisch-semantische Fähigkeiten misst. Wertet man die C-Testergebnisse nach Sprachrichtigkeit und Worterkennung aus, wie Baur & Spettmann es vorschlagen (2006-2009), zeigen sich keine

Differenzen in Bezug auf diese Bereiche. Daher wird bei erwachsenen Muttersprachlern nur eine globale Einschätzung der sprachlichen Fähigkeiten vorgenommen, ohne Fokussierung auf einzelne Bereiche. Lässt der Proband eine oder mehrere Lücken unbearbeitet, so wird davon ausgegangen, dass die Realisierung auf einzelnen Prozessebenen, die z. B. für den Leseprozess zentral wichtig sind, gescheitert ist⁶² (vgl. dazu auch Kap. 5 und 6). Die Rekonstruktion der Lücken könnte aus folgenden Gründen nicht funktioniert haben:

- a) Der Proband kennt die Lösung, schreibt diese aber nicht, da er sich bei der Verschriftlichung (orthografisch oder/und grammatisch) nicht sicher ist.
- b) Der Proband kennt die Lösung nicht und lässt die Lücke daher unausgefüllt.
- c) Der Proband liest den Text schnell durch und füllt nur die Lücken aus, die er sofort rekonstruieren kann, ohne den Text noch einmal zu lesen und die Lösungen nachzutragen.
- d) Der Proband zweifelt die Aussagekraft von C-Tests an und verweigert die Teilnahme an der Testung.

Ziel der vorliegenden Studie war es nicht, eine Analyse dieser Gründe durchzuführen. Die Denkprozesse bei der Bearbeitung von C-Tests haben Feldmann, Grotjahn & Stemmer (1986) und Stemmer (1991) bei Fremdsprachenlernern und anhand von Texten mit der klassischen 2er-Tilgung untersucht. Verwendet wurde dabei die Introspektionsmethode des lauten Denkens. Eines der Hauptergebnisse war, dass C-Testergebnisse von einzelnen Textcharakteristika abzuhängen scheinen; zum Beispiel bestimmt danach die Anzahl der theoretisch möglichen einzufüllenden Elemente die Länge der Entscheidung ebenso wie der Kontextualisierungsgrad des betreffenden Textes: Je mehr Inferenzmöglichkeiten durch den (Kon-)Text gegeben werden, umso leichter und schneller lässt das gesuchte Element sich finden.

Funktionswörter werden mit höherer Erfolgsrate richtig gefunden als Inhaltswörter. Da Funktionswörter weniger zur Bedeutung eines Textes

⁶² Ausführliche Beschreibungen von Vermeidungsstrategien bei der Bearbeitung von C-Tests gibt es bei Klein-Braley (1994) und Grotjahn & Tönshoff (1992).

beitragen, scheint dies die Hypothese zu stützen, dass der C-Test auf einer niedrigen Verarbeitungsebene stattfindet. Die Autorin ist der Meinung, dass das Textverständnis auf einer komplexeren Ebene mit den C-Tests nicht erfolgt. C-Tests sind eine komplexe Aufgabe, die verschiedene Wissenssysteme aktivieren (wenn auch in deutlich unterschiedlichem Ausmaß und Intensität): Textwissen, Sachwissen, Strategiewissen, und zwar in bottom-up-Verarbeitung. Stemmer stellt fest, dass der C-Test in dem von ihr untersuchten Format nicht die allgemeine Sprachkompetenz überprüft, weil der C-Test *low-level-comprehension* auf der maximalen Propositionsebene aktiviert (vgl. dazu Hufeisen, 1997). In der C-Test-Forschung wurde jedoch inzwischen bewiesen, dass C-Tests auch höhere Prozessebenen messen (vgl. Stemmer, 1996). Im Rahmen der vorliegenden Untersuchung wurden die Befunde aus dem fremdsprachlichen Bereich durch den Einsatz von C-Tests bei Muttersprachlern untersucht und erweitert (vgl. dazu Kap. 5). Die Kritik, dass C-Tests das Leseverständnis stärker auf den niedrigeren Prozessebenen messen, ist unter Berücksichtigung der von uns entwickelten Tests und der von uns getesteten muttersprachlichen Adressaten nicht haltbar.

4. Probanden mit Migrationshintergrund werden durch die muttersprachlichen C-Tests benachteiligt, da ihre Ergebnisse an den Ergebnissen der Muttersprachler gemessen und bewertet werden.

Dass die Gruppe der Studierenden, die an der Universität an Veranstaltungen teilnehmen, zu einem hohen Prozentsatz aus Studierenden mit Migrationshintergrund bestehen können, ist eine bekannte Tatsache. Vor allem in Regionen wie dem Ruhrgebiet, das seit den sechziger Jahren zur Einwanderungsregion geworden ist, kommt dies häufiger vor. Ähnlich sieht die Situation an den Schulen aus, wo Klassen mit mehrsprachigen Schülern mit einem Anteil von über 90% keine Seltenheit sind (vgl. Befunde aus der SPREEG-Studie⁶³ an nordrhein-westfälischen Schulen). Die Messung sprachlicher Fähigkeiten von Schülern oder Studierenden mit Migrationshintergrund

⁶³ Zum Projekt SPREEG: <http://www.uni-due.de/daz-daf/spreeg.shtml>, zu den Ergebnissen der Studie vgl. http://www.uni-due.de/unikate/ressourcen/grafiken/PDF%27s/EU_24/24_baur.pdf, zuletzt gesehen am 05.05.2013.

(besonders in gemischten Gruppen) erfolgt in der Regel in Bezug auf die Leistungen der monolingualen Schüler oder Studierenden. Diese Leistungen dienen als Norm, die in dieser Gruppe oder Klasse gültig ist und mit der die Leistungen der anderen Schüler oder Studierenden verglichen werden. Da in der vorliegenden Studie die Sprachkompetenz von Studierenden gemessen wird, wird der Referenznormwert auf der Grundlage einer durchschnittlichen Leistung der monolingualen Studierenden festgelegt. Diese Leistung soll im Durchschnitt auch von Studierenden mit Migrationshintergrund erreicht werden. Würde man die sprachliche Leistung von Studierenden zum Beispiel mit türkischem Migrationshintergrund nur an der durchschnittlichen Leistung türkischsprachiger Studierenden messen, müsste man folgende Variablen erfassen, um eine objektive Bewertung zu ermöglichen: den Geburtsort der Probanden, die Muttersprache der Eltern, die Kommunikationssprache in der Familie, die Aufenthaltsdauer in Deutschland usw. Da an der Universität außerdem über 30 verschiedene Sprachen vorkommen, wäre eine solche Erhebung sehr zeitaufwändig und für die Erfassung der Sprachkompetenz in Deutsch irrelevant. Würde man die Sprachkompetenz Studierender an der Gruppe der gleichsprachigen Studierenden messen, würde man niedrigere Referenzwerte festlegen und die sprachliche ‚Mindestnorm‘ in Deutsch an der Universität nicht erreichen, da die Leistungsbereiche die Entwicklung der sprachlichen Leistungen nicht situations- und Zielgruppen adäquat abbilden würden.

5. Die Textauswahl muss besser bedacht werden. Der Einsatz von fachspezifischen Texten für Lehramtsstudierende setzt Fachwissen voraus und misst daher keine Sprachkompetenz.

Diese Kritik bezieht sich ausschließlich auf die im *Sprachkompetenzprojekt* entwickelten Tests. Die Auswahl der Grundlagentexte für die Studierende beruht auf den Ergebnissen der Erprobungsphase von verschiedenen Texttypen im Projekt *Diagnose & Förderung* an der Universität Duisburg-Essen (vgl. dazu Bremerich-Vos et al., 2013). Getestet wurden folgende Texte: allgemeine, literarische, wissenschaftliche, populärwissenschaftliche und pädagogisch-

propädeutische Texte. Als geeignet haben sich propädeutisch-pädagogische Texte erwiesen. Durch den Einsatz anderer Texte wurde in der Gruppe keine ausreichende Varianz erreicht, da die C-Tests von fast allen Probanden zu über 95 % rekonstruiert wurden. Pädagogisch-propädeutische Texte behandeln Themen, die mit Schule, Bildung und Kindererziehung zu tun haben und sind für zukünftige Lehrer thematisch relevant. Die Grundlagentexte stammen aus den Fachzeitschriften *Psychologie heute* und *Pädagogik*. Der Einsatz dieser Texte lässt sich auch dadurch begründen, dass sie im Sinne von *Testfairness* keine Probanden benachteiligen. In der Regel sind die Texte so geschrieben, dass sie auch für Laien verständlich sind. Mit dem Fokus auf die erste Worthälfte wird u.a. auch das Fachwissen überprüft. Allerdings sind die eingesetzten Texte keine TF-Tests, bei denen gezielt Fachwörter bearbeitet werden (vgl. Kap. 3.5). In den vorliegenden C-Tests wird jedes zweite Wort getilgt. Die Tilgungen können auf unterschiedliche Wortarten fallen, ohne einen bestimmten Fokus auf Fachwörter (z. B. Funktionswörter, Hilfsverben, Präpositionen, Konjunktionen) zu legen. Daher kann der oben genannte Kritikpunkt – auch aufgrund der Testung und späteren Anwendung der Tests – nicht bestätigt werden.

6. Probanden, die das C-Testformat nicht kennen, sind benachteiligt und erzielen niedrigere Werte. Die C-Tests sollen vor der Testung eingeübt werden, so können höhere Werte erzielt werden.

Der dargestellte Kritikpunkt wurde häufiger im schulischen Bereich diskutiert (vgl. zum Beispiel Baur & Mashkovskaya, 2013). Da C-Tests an der Universität Duisburg-Essen von Christine Klein-Braley und Raatz in den achtziger Jahren entwickelt wurden, wird das Thema *Überprüfung der Sprachkompetenz mithilfe von C-Tests* in den Einführungsveranstaltungen DaF/DaZ behandelt, die laut den Studienordnungen an der Universität Duisburg-Essen⁶⁴ mittlerweile von allen Lehramtsstudierenden besucht werden müssen. Vor der Testung wird das Format einmal erklärt. Dass C-Tests trainierbar sind und dadurch bessere C-Test-Werte erzielt werden, konnte in der Forschung bis jetzt nicht bestätigt werden

⁶⁴ http://www.uni-due.de/germanistik/la_lpo2003.shtml, zuletzt gesehen am 17.06.2013.

(vgl. Borgartz & Raatz, 2002). In der Studie wurden Neuntklässler einer Realschule trainiert. Das C-Test-Training verfolgte folgende Ziele:

- a) Verminderung von Unsicherheiten durch möglichst umfassende Informationen und wiederholte Bearbeitung von Übungsmaterial mit eigener Korrekturmöglichkeit.
- b) Verbesserung des Arbeitsverhaltens durch genaue Hinweise zur Instruktion, durch Besprechung von auftretenden Problemen und mehrmalige Bearbeitung von Übungsmaterial.

In der Studie wurde kein signifikanter Leistungsunterschied zwischen der Experimental- und Kontrollgruppe festgestellt. Im schulischen Bereich empfehlen Baur & Spettmann (2007, 2008) den Einsatz von C-Tests als Übungsmaterial. Die Autoren gehen davon aus, dass durch die Arbeit mit C-Tests das Erkennen von Redundanzen und Textbezügen trainiert werden kann, was sich auf bessere C-Testleistungen auswirkt. Derartige Trainings mit C-Tests sind im universitären Bereich nicht bekannt. Vor der Testung könnte auf einzelne Bereiche aufmerksam gemacht werden (z. B. *Beachten Sie Informationen in den tilgungsfreien Sätzen und Überschriften von C-Tests. Hier könnten wichtige Informationen für Sie stehen, auf die sich einzelne Lücken beziehen*). Einzelne C-Tests können jedoch nicht eingeübt werden. Wird ein neuer, für den Probanden unbekannter C-Test eingesetzt, müssen neue Wort-, Satz- und Textbezüge hergestellt werden.

7. Die durch die C-Tests diagnostizierte ‚Risikogruppe‘ stellt immer 25 % der Gesamtgruppe dar. Daher ist die Normierung von C-Tests unnötig.

Sobald neue C-Tests entwickelt werden oder in einer anderen Probandengruppe zum Einsatz kommen, benötigt man Referenzwerte, in deren Bezug die Leistung der Probanden interpretiert werden kann. Liegen Referenzwerte für eine ähnliche Gruppe vor, können sie für die zu testende Gruppe, falls sie der Gruppe in ihrer Beschreibung ähnelt, übernommen werden (z. B. Referenzwerte liegen für Studierende an der Universität X vor, getestet werden aber Studierende gleicher Fachrichtung an der Universität Y). Die Verteilung der

leistungsschwächeren Studierenden in der getesteten Gruppe hängt davon ab, wie schwierig die eingesetzten C-Tests sind und wie das Ziel der Untersuchung lautet. Setzt man schwierige C-Tests ein, differenzieren sie stärker im oberen Bereich unter den Studierenden, die sprachlich leistungsstärker sind und eignen sich z. B. zur Selektion der besten Studierenden. C-Tests mit niedrigem Schwierigkeitsgrad differenzieren stärker im unteren Leistungsbereich und dienen dazu, schwächere Studierende zu identifizieren. Um die Verteilung der leistungsstarken und leistungsschwachen Studierenden in der Gesamtgruppe einzuschätzen, eignen sich C-Tests mit mittlerer Schwierigkeit, die auch in der vorliegenden Studie eingesetzt werden. Wie viel Prozent die sogenannte Risikogruppe ausmacht, hängt davon ab, wo die Referenznorm liegt und ab welcher C-Testleistung ein starker Förderbedarf definiert wird. Da diese Grenzen variierbar sind, kann nicht davon ausgegangen werden, dass die Gruppe der leistungsschwachen Studierenden immer 25 % der Gesamtgruppe ausmacht. Die C-Tests sollten an einer großen Gruppe von Studierenden normiert und skaliert werden, sodass der Referenzwert und die Leistungsgrenzen stabil sind (vgl. dazu ausführlich das Kap. 4.9). Testet man eine leistungsstärkere Gruppe, kann es durchaus vorkommen, dass auf die Gruppe der Studierenden mit hohem Förderbedarf weniger als 10 % der Studierenden entfallen (vgl. Scholten-Akoun & Baur, 2012; Scholten-Akoun, Tischmeyer & Mashkovskaya, 2013). Damit die festgelegten Leistungsbereiche statistisch abgesichert sind und nicht willkürlich gesetzt werden, muss eine Normierung mithilfe der probabilistischen IRT-Modelle erfolgen. Aus den genannten Gründen kann der oben dargestellte Kritikpunkt entkräftet werden.

4.8 Überprüfung von Testgütekriterien von C-Tests für Studierende

Der C-Test genügt in der Regel den Hauptgütekriterien für wissenschaftliche Testverfahren: Reliabilität, Objektivität und Validität⁶⁵, auf die hier noch genauer eingegangen wird. Die für die C-Tests berechneten Reliabilitäten erreichen

⁶⁵ Zum Bereich *Validität* liegen zahlreiche Untersuchungen vor (dazu eine Übersicht über die Validierungsstudien von Grotjahn, 2011), die sich auf C-Tests im fremdsprachlichen Bereich beziehen. Für muttersprachliche C-Tests liegen solche Daten nicht vor.

in der Regel hohe Werte von 0.8 bzw. 0.9 (vgl. Eckes 2007, S. 69). Der C-Test ist objektiv; seine Bearbeitungszeit ist fixierbar und messbar. Die Objektivität der Durchführung wird dadurch gewährleistet, dass die Durchführung nach einer Anleitung erfolgt. Folgende Rahmenbedingungen sollen dabei eingehalten werden:

- Zuerst wird das Testprinzip erklärt, das kann in Form einer Folie oder einer Handlungsbeschreibung mit einigen Beispielen erfolgen oder durch eine mündliche Erklärung.
- Die Tests sollen nacheinander in der vorgegebenen Reihenfolge bearbeitet werden.
- Pro Test werden den Testteilnehmern 5 Minuten zur Verfügung gestellt, mit dem Hinweis, diese Zeit einzuhalten und nicht zum nächsten Test zu blättern ohne vorherige Ankündigung seitens der Prüfer (vgl. Baur & Spettmann, 2007, S. 126; Baur & Spettmann 2008, 433).

Die Instruktionen für die zeitlimitierte C-Test-Variante⁶⁶ sind bei Wockenfuß (2009) und Grotjahn, Schlak & Aguado (2010), Grotjahn (2010b) ausführlich beschrieben.

Um Objektivität der Durchführung zu gewährleisten, können parallele Test-Varianten verwendet werden, z. B. in engen Räumlichkeiten oder mit zu vielen Probanden. Dadurch kann vermieden werden, dass die Testteilnehmer voneinander abschreiben, was die Testergebnisse verfälschen würde.

Die Objektivität der Auswertung ist durch die Antwortschablonen gegeben: Die im Originaltext vorgegebenen Lösungen werden mit den Lösungen der Testteilnehmer abgeglichen; zusätzlich werden alternative Lösungsvarianten erfasst, die dem Text kontext- und sinngemäß entsprechen.

Dem Kriterium Validität widmen sich mehrere Untersuchungen (vgl. u. a. Grotjahn, Klein-Braley & Raatz 2002; Arras, Eckes & Grotjahn 2002; Eckes & Grotjahn 2006a,b Grotjahn 2010a). Sie gehen der Frage nach, was genau die C-Tests messen. Klein-Braley (1994) untersucht in ihrer Habilitationsschrift genau diesen Punkt und hat herausgefunden, dass C-Tests besonders für die

⁶⁶ Die zeitlimitierten C-Tests wurden zur Überprüfung der Intelligenz und des Bearbeitungstempos von Studierenden untersucht (vgl. Wockenfuß 2009, S. A31f.).

Überprüfung der allgemeinsprachlichen Kompetenz von L2-Lernern geeignet sind. C-Tests korrelieren danach bei fortgeschrittenen Lernern des Englischen als Fremdsprache mit unterschiedlichen Testverfahren, die ebenso Aspekte der sprachlichen Kompetenz überprüfen (Schreiben, Lesen, Sprechen und Hören). Diese Ergebnisse sind allerdings nicht für Zweitsprachenlerner und Muttersprachler überprüft worden; hier könnten durchaus andere Korrelationen bestehen.

Validitätsstudien verlaufen in der Regel mit der Messung eines sprachlichen Merkmals, das von einer zusätzlichen Testaufgabe erfasst wird. Beide Testwerte werden miteinander korreliert, um eventuelle Zusammenhänge zu beweisen. Das wurde bei den C-Tests mit der Erfassung der Lehrerurteile, Deutschnoten, Selbsteinschätzung der Kompetenz in der deutschen Sprache und anderen Eigenschaften erfasst.

Die C-Test-Ergebnisse können sowohl mit der klassischen als auch mit der probabilistischen Testtheorie analysiert werden. Die Interpretation der Testergebnisse ist von der Gültigkeit des ausgewählten Modells abhängig (zum Beispiel Anwendung der Rasch-Modelle)⁶⁷.

Welche der beiden Testtheorien in Frage kommt, ist von der Auswertung, Interpretation der Testergebnisse und dem Zweck der Untersuchung abhängig. Bei der Anwendung der klassischen Testtheorie muss beachtet werden, dass die Testwerte mit einem Messfehler behaftet und die Daten stichprobenabhängig sind. Die probabilistische Testtheorie erlaubt es, die Itemeigenschaften und Personeneigenschaften auf einer Logit-Skala zu positionieren, d. h. die Testwerte werden in Bezug auf die Aufgabenschwierigkeit hin interpretiert. Erreicht die Person X einen hohen Wert bei einer leichten Aufgabe, wird er anders interpretiert als der Testwert einer eher schwierigeren Aufgabe.

Bei der Überprüfung der psychometrischen Kriterien sind folgende wichtige Eigenschaften zu untersuchen: Reliabilität, Kontextvalidität, kognitive Validität, Validität der Ergebnisermittlung, Textwirkungsvalidität und kriterienbezogene Validität⁶⁸. So kann die Frage beantwortet werden, ob die Testergebnisse

⁶⁷ Analysen zur Überprüfung der Geltung von Rasch-Modellen bei C-Tests finden sich bei Eckes & Grotjahn; Eckes, 2006, 2007, 2010b, d, 2011.

⁶⁸ Mehr zum Thema Konstruktvalidität bei Khalifa & Weir, 2009.

zuverlässig sind und der Test wirklich das Merkmal misst, das er messen soll. Psychometrische Analysen zum C-Test werden detaillierter im Kap. 10 dargestellt. Außer den genannten Kriterien heben Lienert und Raatz hervor, dass der C-Test normierbar⁶⁹, vergleichbar⁷⁰ und nützlich⁷¹ sei (Lienert & Raatz 1998, S. 8 ff.).

4.9 Auswertung und Interpretation von C-Testergebnissen

Für die Testauswertung müssen strenge Auswertungsregeln festgelegt werden, um das Kriterium der Objektivität zu erfüllen. Die wichtigsten Festlegungen sind: a) welche Lücken werden als richtig bzw. falsch gewertet? b) Wie wird mit den unterschiedlichen Varianten umgegangen? c) Wie werden nicht rekonstruierte Lücken gewertet?

In Abhängigkeit von den Auswertungsregeln ergeben sich bestimmte Codierschemata. In unserer Untersuchung erhält jede dem Original entsprechende Lösung einen Punkt. Jede falsch oder nicht rekonstruierte Lücke wird mit keinem Punkt bewertet. Morphosyntaktisch und semantisch angemessene Alternativen werden anhand vorher erstellter Lösungslisten als korrekt gewertet⁷².

Die Summe der korrekten Lösungen eines C-Tests ergibt den Score für den einzelnen *Teilttest*⁷³ und die Summe der Punktwerte der Teilttests den Gesamtwert für die Testbatterie. Der *Gesamtwert* oder *Gesamtscore* liefert erste Informationen über die globale Sprachfähigkeit der Testteilnehmer, ohne weitere Hinweise auf Problemfelder der Probanden.

⁶⁹ Normwerte für C-Test für die Klassenstufen 4-8 aller Schultypen in NRW werden von Baur & Spettmann erhoben.

⁷⁰ Die Leistungen der Testteilnehmer können untereinander verglichen werden (zum Beispiel Bildung von Rangreihen). Die Interpretation der Abstände zwischen den Punktwerten ist mit dem C-Test nicht möglich.

⁷¹ Das Kriterium der Nützlichkeit beschreibt die Fragen der Sprachverwendung in der realen Situation.

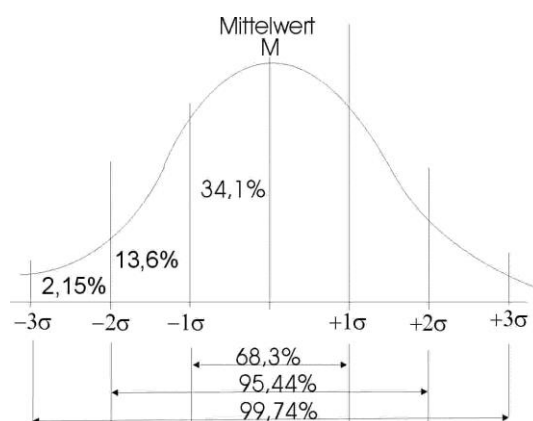
⁷² Es gibt verschiedene Varianten und Alternativen, wie man C-Test-Lücken erfasst und analysiert, z. B. können orthografisch falsch, aber sinngemäß richtig rekonstruierte Lücken separat erfasst werden, das hängt stark vom Ziel der Untersuchung ab (vgl. auch Untersuchungen von Baur & Spettmann oder Grotjahn, 2002).

⁷³ Die Teilttests werden in unserer Untersuchung als Items bezeichnet.

Jeder Testteilnehmer wird aufgrund seines Testergebnisses normativ einer Leistungsgruppe zugeordnet. Auf diese Weise erfolgt ein *Gruppen-Ranking*. Es basiert auf den summierten Testergebnissen in den einzelnen Teilttests⁷⁴. Auf diese Weise wird jedem Testteilnehmer innerhalb der getesteten Gruppe eine relative Position zugewiesen. Das Testergebnis wird darüber hinaus in Relation zum Gruppendurchschnitt (Gruppen-Normwert) gesetzt. *Der Gruppen-Normwert* oder *Referenz-Normwert* wird an den Testergebnissen der Testteilnehmer mit der Muttersprache Deutsch ermittelt. Die Leistungen der bilingualen Studierenden werden dabei nicht beachtet⁷⁵. Wir betonen die Wichtigkeit der Messung der Sprachkompetenz an der Norm der Muttersprachler. Es wäre unseres Erachtens nach problematisch, die Berechnung der Norm an Studierende mit Migrationshintergrund zu koppeln, weil dadurch zum einen andere Referenz-Normwerte gebildet werden könnten und zum anderen sich der Abstand zwischen den Leistungsgruppen eventuell vergrößern würde (vgl. auch das Kapitel zu der Kritik von C-Tests).

Bei der Bildung der Leistungsgruppen wird als Maß die Standardabweichung herangezogen (dazu Schmidt-Atzert & Amelang, 2006, S. 162). Hier wird beispielhaft die Anwendung dieser Methode für die C-Test-Daten gezeigt.

Abb. 6: Beispiel für die Normierung der Daten



aus: Schmidt-Atzert & Amelang, 2006, S. 162

Anmerkung: Als σ (Sigma) wird in der Abb. 1 die Standardabweichung gekennzeichnet.

⁷⁴ $T1 + T2 + T3 + T4 = \text{Gesamtscore}$.

⁷⁵ Die Scores der Studierenden mit Migrationshintergrund werden bei der Berechnung des Referenz-Normwertes herausgerechnet.

Durch die Gruppierung der Testwerte um den Mittelwert (eine Standardabweichung oberhalb und unterhalb des Normwertes) können 68,3 % der Fälle beschrieben werden. Durch die Ausweitung der Standardabweichungen um das Zweifache oberhalb und unterhalb der Normwerte können 95,44 % der Untersuchungsfälle erklärt werden. Bei einer weiteren Berücksichtigung der Bereiche handelt es sich um Extremfälle oder Ausreißer im oberen und unteren Bereich. Die Differenzierung der Leistungsgruppen wird bei Baur & Spettmann für den schulischen Bereich beschrieben (Baur & Spettmann 2007, S. 101; Baur & Spettmann 2010, S. 436). Die Normwerte für die Beschreibung der sprachlichen Fähigkeiten bei erwachsenen Muttersprachlern im universitären Bereich wurden in der Forschung bislang wenig untersucht und tauchen zuerst bei Bremerich-Vos et al. (2013) auf. Die Heranziehung der Standardabweichung als Maß für die Festlegung der Leistungsgruppe erscheint in diesem Zusammenhang sinnvoll. Die Normierung der C-Tests wird aus dem *Sprachkompetenzprojekt Diagnose & Förderung* übernommen. Für die Festlegung des Referenznormwerts wurden 2002 Studierende an nordrhein-westfälischen Universitäten getestet.

Ab einer Unterschreitung des Normwertes um zwei Standardabweichungen wird ein Förderbedarf unterstellt. Welche sprachlichen Fertigkeiten dabei als förderbedürftig gelten, konnte über C-Tests bis jetzt nicht ermittelt werden. Die C-Tests erfassen, wie bereits diskutiert, nur die allgemeine Sprachkompetenz – eine Auswertung nach Kriterien der formalsprachlichen Richtigkeit (orthografische, semantische und grammatische) und des allgemeinen Textverständnisses (Rekonstruktion der Lücken, die den Sinn des Textes nicht verfälschen) ist nur eingeschränkt möglich. Die Ergebnisse des C-Tests dienen nur als Indikator für sprachliche Defizite, die im zweiten Schritt über geeignete Testaufgaben genauer identifiziert werden müssen (vgl. Scholten-Akoun & Baur, 2010).

Tabelle 5: Beispiel für die Festlegung der Leistungsgruppen

Stufen	Zuordnung zu den Leistungsbereichen
1	> 2 Standardabweichungen über dem Referenznormwert
2	1 bis 2 Standardabweichungen über dem Referenznormwert
3	\pm 1 Standardabweichung um den Referenznormwert
4	1 bis 2 Standardabweichungen unter dem Referenznormwert
5	> 2 Standardabweichungen unter dem Referenznormwert

Die Testergebnisse der TestteilnehmerInnen können auf der Grundlage der oben aufgeführten Tabelle nur normativ interpretiert werden. Die sprachlichen Leistungen der Probanden, die den **Gruppen 1 und 2** zugeordnet werden, liegt 1 bis 2 Standardabweichungen über dem Gruppen-Normwert. Auf sprachliche Förderung kann bei den Testteilnehmern dieser Gruppe sehr wahrscheinlich verzichtet werden, da die C-Testergebnisse sprachlich nicht auffällig sind. Die Testergebnisse sind laut dem für den C-Test für die Gesamtgruppe errechneten Normwert dem obersten Kompetenzbereich zuzuordnen, und es wird angenommen, dass diese Personen auf verschiedenen Ebenen (z. B. syntaktische und semantische Ebene, Textsortenwissen, Texterschließungsstrategien etc.) über eine gut ausgebildete allgemeine Sprachkompetenz verfügen.

In der **Gruppe 3** werden die Testergebnisse derjenigen Testteilnehmer erfasst, deren Testergebnisse im mittleren Leistungsbereich liegen und 1 Standardabweichung über und unter dem Normwert abweichen. Die Leistungen dieser Gruppe werden als akzeptabel interpretiert. Es gibt keine großen sprachlichen Auffälligkeiten, die durch C-Tests identifiziert werden können.

Den Personen der **Gruppen 4 und 5** wird sprachlicher Förderbedarf unterstellt; diese Gruppe wird als *Risikogruppe* bezeichnet. Mehrere fehlerhafte Lösungen treten bei der Rekonstruktion der C-Tests auf. Die Fehler sind unterschiedlichen Kategorien zuzuordnen. In welchen Bereichen die sprachlichen Schwierigkeiten liegen, kann an dieser Stelle mit dem C-Test nicht diagnostiziert werden. Dafür sollen andere diagnostische Testverfahren verwendet werden.

Mit der vorliegenden Studie wird versucht, über die globale Einschätzung der Sprachkompetenz hinauszugehen und die Lesefähigkeit der Studierenden durch

den Einsatz von C-Tests zu bewerten. Wie der C-Test in Bezug auf die Lesefähigkeiten auszuwerten und zu interpretieren ist, wird in Kap. 6 diskutiert.

5. Lesetests

Lesekompetenz besteht aus einer Vielzahl von Merkmalen und kann durch geeignete Lesetests erfasst werden, die unterschiedliche Teilkompetenzen messen, z. B. Textverständnis oder Lesegeschwindigkeit. Es gibt eine große Anzahl von Lesetests, die für unterschiedliche Zwecke eingesetzt werden. Die Auswahl der Testverfahren zur Überprüfung der Lesekompetenz erfordert grundlegende Überlegungen. Zum einen ist die Beschreibung der Zielgruppe entscheidend, denn von ihr hängt sowohl die Auswahl des Testkonstrukts ab als auch (zum Teil) das Ziel der Testung. Zum anderen sind das Testformat, der Testinhalt und das anzuwendende Bewertungsverfahren wichtig.

Im Folgenden werden verschiedene Lesetests dargestellt, die sowohl nach Testformat als auch nach Zielgruppe unterschieden werden. Es werden Testverfahren skizziert, die sich für die Anwendung mit Muttersprachlern (Deutsch als Muttersprache) und Fremdsprachenlernern (Deutsch als Fremdsprache) eignen. Bevor auf unterschiedliche Zielgruppen eingegangen wird, werden Ziele von Lesetests, Lesetestformate und grundlegende Fragen vor dem Einsatz von Lesetests diskutiert.

5.1 Lesetestformate und grundlegende Fragen vor dem Einsatz von Lesetests

Vom Ziel der Testung und der zu testenden Gruppe hängt auch die Auswahl des Testformats ab. Für eine inhaltliche Überprüfung der Zusammenfassung eines Lesetextes eignen sich gut *offene* Aufgabentypen. Gibt man den Probanden lediglich eine Information, ohne Antwortmöglichkeiten zu nennen und ohne eine bestimmte Antwort zu erwarten, handelt es sich um eine offene Aufgabe. Offene Aufgaben kommen zum Einsatz, wenn man das Zusammenspiel von Kompetenzen verschiedener Wissensgebieten in Bezug auf Problemlösungen trainieren bzw. testen will. Das selbstständige Agieren in Verbindung mit kreativem Denken steht dabei im Vordergrund.

Will man das Textverständnis überprüfen, indem man die Testteilnehmenden kurze Fragen zum Text beantworten lässt, sind *geschlossene* oder *halboffene* Aufgabenformate mit Ja/Nein- oder Richtig/Falsch-Antwortoptionen besser geeignet (vgl. dazu Bolten, 1996, S. 26 ff.).

Bei *geschlossenen Aufgaben* handelt es sich um Entscheidungsaufgaben, die von zwei vorgegebenen Antworten nur eine als richtig gelten lassen. Zu den geschlossenen Aufgaben zählen folgende Formate: *Richtig/Falsch-Aufgaben*, *Multiple-Choice-Aufgaben*, die Mehrfachantworten anbieten, *Zuordnungsaufgaben*, die einen oder mehrere Lösungsblöcke vorgeben und die eine korrekte Zuordnung verlangen (z. B. Lückentexte mit Antwortvorgaben).

Der Aufgabentyp der *halboffenen Aufgaben* zeichnet sich dadurch aus, dass Probanden selbstständig mit eigenen Worten Lösungen präsentieren müssen. Die Bandbreite der Antworten kann dabei von nur einem Wort über Wortgruppen und Sätzen bis hin zu zusammenhängenden Texten unterschiedlicher Länge reichen. Die Steuerung einer möglichen Lösung übernimmt hierbei die konkrete Aufgabenformulierung mittels Aufgabenform. Die Aufgabe kann zum Beispiel auffordern, Informationen aus einem Text zu ermitteln, diese zu ordnen, unter vorgegebenen Aspekten zusammenzustellen, zu vergleichen usw.

In standardisierten Tests werden häufig halboffene oder geschlossene Testformate verwendet, z. B. in Form von Multiple-Choice- oder Zuordnungsaufgaben. In nicht standardisierten Tests, besonders bei der Überprüfung der Lesekompetenz im schulischen Unterricht, werden eher offene Aufgaben eingesetzt. Der Einsatz offener Aufgaben ist nicht unproblematisch, da solche Aufgabenformate in der Regel auch ein Element des Schreibens enthalten, sodass Lese- und Schreibkompetenz nicht immer völlig eindeutig voneinander zu trennen sind. Für die Bewertung muss für den Test ein Raster mit möglichen Antwortoptionen erstellt werden, was für die Praxis in Lehr-Lern-Institutionen aus Gründen der Ökonomie nicht immer praktikabel ist (zu weiteren Schwierigkeiten vgl. ALTE, Modul 3, S. 22f.).

Bei der Entscheidung für oder gegen ein bestimmtes Aufgabenformat von Lesetests werden vor der Anwendung im Folgenden in Anlehnung an Arras (2006, S. 86) relevante Fragen dargestellt, die im Wesentlichen die Zielgruppe, das Testkonstrukt, die Textauswahl und die Bewertung betreffen:

- **Zielgruppe:** Wer soll geprüft werden?
Beschreibung der Zielgruppe (homogen bzw. heterogen), Alter, kultureller Hintergrund, Lernbiografie.
- **Testkonstrukt & Textauswahl:** Was soll geprüft werden?
Welche Items (Fragen) sind hinsichtlich der zu überprüfenden Fähigkeiten angemessen?
Wie können das Ergebnis verfälschende Teststrategien (z. B. *test guessing*) kontrolliert werden?

Welche Textsorten werden dem Testkonstrukt zugrunde gelegt?
Welche Merkmale soll der Lesetext aufweisen? Wie vertraut sind die Prüflinge mit der Textsorte?
Beschreibung des Textes (Textsorte, Textlänge, sprachliche Strukturen, Merkmale etc.). Ist das Thema für die Prüflinge bekannt? Wie viel Vorwissen ist erforderlich?
Wie viel strategisches/kulturelles Wissen ist für die Lösung der Aufgabe erforderlich?
- **Bewertung:** Wie kann eine zuverlässige Bewertung erreicht werden?
Wie praktikabel muss das Auswertungsverfahren sein (geschlossene, halboffene, offene Aufgabentypen)? Welche Ressourcen stehen zur Verfügung?

In Bezug auf die getestete Gruppe in der vorliegenden Studie lassen diese Fragen sich folgendermaßen beantworten.

Testgruppe:

Die Testgruppe besteht, wie bereits oben erwähnt, aus Studierenden, die den Lehrerberuf anstreben und sich hauptsächlich im Grundstudium befinden. Das Alter der Studierenden liegt durchschnittlich zwischen 21-23 Jahren. Etwa zwei Drittel der Studierenden sprechen Deutsch als Muttersprache, ein Drittel ist in mehrsprachigen Familiensituationen aufgewachsen, d. h. in der Familie wurde außer Deutsch noch eine andere Sprache gesprochen. Die Probanden verfügen über ein deutsches Abitur oder eine andere Prüfung, die zur Aufnahme eines Studiums an einer deutschen Universität berechtigt. Kurz zusammengefasst: Die

Gruppe ist in Bezug auf das Alter und den angestrebten universitären Abschluss homogen, in Bezug auf den kulturellen Hintergrund jedoch eher heterogen.

Testkonstrukt & Textauswahl:

Da im Studium das Lesen eine Grundlage für die erfolgreiche Entwicklung von Schreibfähigkeiten auf akademischem Niveau bildet (z. B. Erkennen und Beherrschen wissenschaftlicher Strukturen, Verstehen von Pro- und Contra-Argumentation in einem wissenschaftlichen Text), sollen die Studierenden die oben genannten Strukturen in einem Text erkennen. In den Lesetests soll das Verstehen von Textinhalten für Texte überprüft werden, die für die Zielgruppe als potenziell studienrelevant angesehen werden können (zur Auswahl der entsprechenden Texte s. Kap. 5.4.1, Kap. 5.4.2) Im Einzelnen betrifft die Aufgabenstellung die Entnahme von wichtigen Informationen, globales, gezieltes und detailliertes Verstehen von Informationen sowie Verstehen von impliziten Informationen. Die Lesetests müssen mit einer zeitlichen Beschränkung (unter Zeitdruck) gelöst werden, was dem Gütekriterium der Authentizität zwischen Test- und Anwendungssituation entspricht. Können Studierende schneller lesen und dabei verschiedene Lesetechniken anwenden (z. B. gezielt nach relevanten Informationen im Text suchen), werden sie im Studium mit Leseaufgaben vermutlich besser zurechtkommen.

Als Grundlage für die Lesetests sollen längere (ca. 500-600 Wörter) allgemeine, populärwissenschaftliche Texte aus einer Zeitung oder einer Zeitschrift genutzt werden, die ein wissenschaftliches Thema behandeln, das von einem Laienpublikum verstanden werden kann. Der Text soll einen klaren Aufbau besitzen, aber auch komplexe sprachliche Strukturen enthalten (z. B. Wortbildung, wissenschaftssprachliche Sprache, Nominalisierungen, komplexe Satzstrukturen). Er soll eine hohe Informationsdichte aufweisen und auch über abstrakte Informationen verfügen.

Für das Verständnis der Texte und die Beantwortung der Fragen sollten die im Text vorhandenen Informationen ausreichen. Spezialwissen sollte nicht erforderlich sein. Durch die Entnahme wichtiger inhaltlicher Informationen aus dem propädeutisch-pädagogischen Kontext und durch gezieltes und detailliertes

Lesen soll die Fähigkeit des Umgangs mit den im universitären Lehramtsstudium vorkommenden Lesetexten geprüft werden. Für die Suche nach geeigneten Informationen sollten die Probanden verschiedene Lesestrategien aktivieren müssen. Um 'test guessing' zu vermeiden, soll ein für die getestete Gruppe geeignetes Testformat verwendet werden (z. B. Multiple-Choice-Lesetests mit erweiterten Antwortoptionen).

Bewertung:

Für die Bewertung der Lesetests sollten wegen der Ökonomie, der personellen und zeitlichen Ressourcen und der Objektivität der Auswertung geschlossene Aufgabentypen verwendet werden (z. B. Multiple-Choice-Lesetests oder Zuordnungsaufgaben).

Da es in meiner Studie darum geht, den C-Test zu überprüfen, der im Rahmen der Sprachkompetenzprojekte Diagnose & Förderung als Messinstrument für die Lesefertigkeit genutzt wird, wird an dieser Stelle auf den C-Test nicht weiter eingegangen (vgl. dazu Kap. 4 und Kap. 7).

Die von Arras (2006) formulierten Fragen vor dem Einsatz von Lesetests können durch weitere Kriterien vervollständigt werden, die für die Entscheidung für oder gegen ein Testformat von Bedeutung sind:

Für die Auswahl eines geeigneten Lesetests sollte das Kriterium *Testnützlichkeit* überprüft werden. Dieses Testgütekriterium lässt sich unter dem von Arras formulierten Kriterium Textauswahl/-konstrukt subsummieren. Bachman & Palmer (1996, S. 18) definieren Testnützlichkeit als Funktion von sechs komplementären Eigenschaften:

Testnützlichkeit=

Reliabilität+Konstruktvalidität+Authentizität +Interaktivität+Impact+Praktikabilität

Im Folgenden werden die zwei Eigenschaften *Authentizität* und *Praktikabilität* kommentiert, da sie für die Auswahl des Testformats entscheidend sind. Die Beschreibung anderer Eigenschaften, die in Bezug auf Bereiche wie z. B. Überprüfung des Testkonstrukts oder Auswirkungen auf den Unterricht (Impact) wichtig sind, werden bei Bachman & Palmer (1996) und Grotjahn (2003) beschrieben.

Authentizität bezieht sich auf genuine, d. h. authentische, nicht spezifisch für den Test konstruierte Aufgaben. Authentizität ist für die Testung aus zwei Gründen wichtig (zur Auswahl authentischer Materialien zur Überprüfung des Leseverstehens vgl. Bolten, S. 21):

- 1) Authentische Aufgaben erlauben eine Generalisierung im Hinblick auf die Fähigkeit zur Lösung analoger sprachlicher Probleme außerhalb der Testsituation,
- 2) Authentizität bestimmt die Wahrnehmung der Aufgabe durch den jeweiligen Probanden als situationsrelevant und im Sinne der Augenscheinvalidität als situationsangemessen (vgl. Grotjahn, 2003, S. 28f.).

Praktikabilität wird als das Vorhandensein angemessener Testbedingungen verstanden, die Einfluss auf die Textauswahl und die Testdurchführung haben. Sie kann mit dem Kriterium der Ökonomie in Konflikt geraten. Zum Beispiel werden bei Testungen aus Gründen der Praktikabilität (Ökonomie) öfter kürzere Texte verwendet. Durch den Verzicht auf längere Texte werden jedoch im akademischen Bereich wichtige Fähigkeiten, die z. B. zum Erfassen eines argumentativen Textes notwendig sind, nicht adäquat gemessen (vgl. Grotjahn, 2003, S. 90).

Eine geringe Praktikabilität kann zum Beispiel aufgrund sehr komplexer Durchführungs- und Auswertungsverfahren unter anderem zu Messfehlern führen und die Reliabilität und Validität herabsetzen (vgl. Grotjahn, 2003, S. 160). Praktikabilität hat in diesem Zusammenhang also sowohl für die Textauswahl und -konstruktion als auch für die Bewertung Konsequenzen.

Im Folgenden wird dargestellt, wie die Kriterien Authentizität, Praktikabilität, Reliabilität und Konstruktvalidität in Bezug auf die getestete Gruppe zu beschreiben und umzusetzen sind.

Überprüfung der Kriterien Authentizität und Praktikabilität

Für die Überprüfung der Lesekompetenz der Studierenden sollen authentische, nicht spezifisch für den Test konstruierte Aufgaben verwendet werden. Für diesen Zweck sollen Texte gewählt werden, die nicht nur für den Hochschulkontext relevant sind (z. B. Bildung, Erziehung, Psychologie, Pädagogik), sondern auch eine spezifische Thematik haben, die mit Bildung, Lernen, Schule und Gesellschaft verbunden ist, da die Studierenden den Lehrerberuf anstreben.

Für die Testung kommen nur solche Lesetests in Frage, die in Bezug auf Durchführung und Auswertung ökonomisch sind. Für die Lesetests sollen sowohl kurze als auch längere Texte verwendet werden, um die Lesefähigkeiten in Bezug auf zwei verschiedene Textsorten zu prüfen. Ein kurzer und ein langer Text, die unterschiedliche Aufgabenstellungen enthalten, können aufgrund unterschiedlicher Schwierigkeitsgrade die Leseleistung der Studierenden differenzieren, sodass zwischen leistungsstarken und leistungsschwachen Lesern unterschieden werden kann. Ein kurzer Text aus der „Kritik der praktischen Vernunft“ von Kant zum Thema Beziehung und ein langer Text aus GEO-Reisen-Venedig weisen zum Beispiel einen unterschiedlichen Grad an Informationsdichte auf. Scheitert ein Leser aufgrund seiner Arbeitsgedächtniskapazität beim Lösen der schwierigeren Leseaufgabe, kann er dieses Leistungsdefizit bei einer weiteren Leseaufgabe, die kürzer ist, kompensieren.

Erweist sich einer der vorgetesteten Lesetests in Bezug auf eines der Kriterien als ungeeignet (etwa, wenn die Reliabilität des einzusetzenden Lesetests für die getestete Gruppe nicht ausreichend ist), wird er aus der Testung herausgenommen. Falls Lesetests aus anderen Untersuchungen oder Studien übernommen werden, sollen entsprechende Reliabilitäts- und Validitätswerte analysiert werden. Wird die Validität auf einem korrelativen Weg beschrieben, so sollen die Korrelationen mit anderen Testverfahren mindestens mittelhoch sein.

5.2 Überblick über Lesetests für Deutsch als Muttersprache (DaM)

Die Anzahl der auf dem Markt vorhandenen Lesetests ist groß und wird fast jährlich um weitere, aktualisierte Neuauflagen oder um neue Testverfahren ergänzt. Ein umfassender Überblick über Lesetests für verschiedene Klassenstufen findet sich z. B. bei Lenhard & Schneider (2009) und Souvignier (2009).

Obwohl das Lesen als ein hochkomplexer Prozess beschrieben wird und es schwierig ist, die dabei ablaufenden Prozesse diagnostisch zu erfassen, gibt es dennoch Lesetests, die versuchen, hinreichend messgenaue Aussagen über einzelne Teilleseleistungen zu erreichen. Hierbei ist es oft Ziel der Testung, in einem in Bezug auf den Zeitaufwand ökonomischen zehn- bis fünfzehnminütigen Screening-Verfahren auf Leseschwierigkeiten zu schließen, die dann aber diagnostisch erst mit differenzierteren Tests (ca. 30-45 Minuten) präzisiert werden können.

In Anlehnung an die Kategorisierung der Testverfahren zur Erfassung der Lesekompetenz von Souvignier (2009) werden drei Testverfahrensgruppen dargestellt:

- (1) Testverfahren, die Basiskompetenzen erfassen (z. B. Lesegeschwindigkeit)
- (2) Testverfahren, die komplexere Leseverständnisleistungen erfassen
- (3) Testverfahren, die das Leseverständnis und seine Bedingungsfaktoren erfassen

Bei der ersten Kategorie handelt sich um ein Screeningverfahren auf Satzebene, das Hinweise über eventuell vorhandenen Förderbedarf im Bereich der Lesegeschwindigkeit gibt. Ein Beispiel dafür wären das *Salzburger Lese-Screening für die Klassenstufen 1-4* (abgekürzt SLS 1-4) oder das *Salzburger Lese-Screening für die Klassenstufen 5-8* (SLS 5-8). Überprüft wird hier, ob die Lesenden in begrenzter Zeit für möglichst viele einfache Aussagen entscheiden können, ob sie richtig sind oder falsch. Diese Verfahren geben vor allem Auskunft über die Lesegeschwindigkeit. Keine Auskunft geben sie darüber, ob langsam und sorgfältig oder eher schnell und ungenau gelesen wird.

In Bezug auf die zu testende Gruppe der Studierenden lässt sich die Basiskompetenz Lesegeschwindigkeit ebenso erfassen. Allerdings sollten diese Leseaufgaben inhaltlich an das Niveau der Studierenden angepasst werden. In der Leseforschung ist ein für Grundschüler entwickelter Lesetest bekannt, der von Brügelmann und Backhaus (2004) an Studierenden und Lehrern getestet wurde, ohne an die Gruppe der Lehrer inhaltlich angepasst zu werden. Der Lesetest trägt den Namen *Stolperwörterlesetest*⁷⁶. Stolperwörter-Tests prüfen, wie schnell und genau jemand liest und was er dabei versteht (vgl. dazu Kap. 5.4.2).

Bei der zweiten Kategorie handelt es sich um Tests, bei denen ganze Texte bearbeitet werden. Dabei werden Fragen zum Zusammenhang zwischen verschiedenen Textteilen gestellt. Für die Bearbeitung dieser Tests ist ein Gesamtverständnis des Textes erforderlich. Ein Beispiel dafür ist der *Frankfurter Leseverständnistest 5-6* (FLVT 5-6) – ein Testverfahren zur Erfassung des Leseverständnisses in der 5. und 6. Klasse. Er besteht aus einem Leseverständnistest zu einer Geschichte und einem Leseverständnistest zu einem Sachtext. Zu beiden Texten (ca. 560 Wörter) sind je 18 Fragen im Multiple-Choice-Format zu beantworten. Anhand der Ergebnisse können Schülerinnen und Schüler einem bestimmten Fertigniveau des sinnentnehmenden Lesens zugeordnet werden. Allerdings erlaubt der Test keine Aussagen darüber, welche Bedingungsfaktoren (Motivation, Lesestrategien oder Wortschatz) für das erzielte Ergebnis ausschlaggebend sind.

Der Einsatz solcher Lesetests ist auch für die Gruppe der Studierenden denkbar, weil sie auf ökonomische Weise sinnentnehmendes Lesen überprüfen. Was die Textauswahl für ein solches Testformat betrifft, sollten studienrelevante Themen gewählt und an das Niveau der zu testenden Gruppe sprachlich angepasst werden.

Die dritte Kategorie bilden Testverfahren, die das Leseverständnis und seine Bedingungsfaktoren erfassen. Die Kategorie beinhaltet folgende inhaltliche

⁷⁶ Online-Artikel: Lese-Test. Schüler kaum schlechter als ihre Lehrer, unter: <http://www.spiegel.de/schulspiegel/lese-test-schueler-kaum-schlechter-als-ihre-lehrer-a-286665.html>, zuletzt gesehen am 19.06.2013. Ein Ergebnis der Studie ist, dass ein Drittel der zehnjährigen Schüler die Genauigkeit, Geschwindigkeit und das inhaltliche Verständnis im Lesetest erreicht wie als leseschwach eingeschätzte LehrerInnen.

Bereiche: Leseverständnis, Lesestrategiewissen und Lesemotivation. In den Bereichen Überprüfung des Lesestrategiewissens und Lesemotivation liegen kaum verfügbare standardisierte Tests vor. Daher ist es notwendig, Leseverständnistests mit ergänzenden Verfahren zu kombinieren. Ein Beispiel für einen Lesestrategietest stellt *der Würzburger Lesestrategie-Wissenstest für die Jahrgangsstufen 7-12 (WLS)* dar. Bei ihm sind von den Schülerinnen und Schülern Strategievorschläge für verschiedene Lesesituationen mit Schulnoten zu bewerten. Der WLST dient der Erfassung des Lesestrategiewissens von Schülern der Klassenstufen 7 bis 12. Den Schülern werden sechs verschiedene Lernszenarien dargeboten. Zu jeder Situation sollen sie die Qualität und Nützlichkeit fünf verschiedener Vorgehensweisen zur Erreichung eines Lernziels bewerten. Dazu werden verschiedene Antwortvorschläge vorgegeben, welche die Schüler hinsichtlich ihrer Qualität benoten sollen. Hohe Testwerte zeigen, dass der Schüler über effektive Strategien zum Behalten und Verstehen von gelesenen Texten verfügt. Der Test kann als Gruppen- und Einzeltest durchgeführt werden; seine Bearbeitungszeit beträgt zwischen 20 und 35 Minuten. Zur Erfassung des lesebezogenen Selbstkonzepts stehen Bewertungsskalen mit folgenden Antworten zur Verfügung, z. B. *Es fällt mir leicht, neue Texte zu verstehen*. In ähnlicher Weise lässt sich auch das Leseinteresse erfragen, z. B. *Ich lese gerne in meiner Freizeit*.

Der Einsatz von dieser Kategorie von Lesetests wäre im universitären Bereich problematisch. Zum einen sind solche Tests nicht standardisiert, und zum anderen erfordern sie den Einsatz kombinierter Leseverständnistests. Diese Vorgehensweise würde den Kriterien Testnützlichkeit und Praktikabilität nicht entsprechen.

Nachfolgend wird in tabellarischer Form ein Überblick über einige aktuelle deutschsprachige Lesetests gegeben. Neben den Testnamen, den Abkürzungen und dem Autor bzw. den Autoren wird das Jahr der Anwendung (der aktuellen Auflage), die Zielgruppe (Alter oder Schulklasse) bzw. der Einsatzbereich und der Messgegenstand sowie Angaben zur Testnormierung genannt. Die vorhandenen Lesetests sollen u.a. auch in Bezug auf die ersten zwei Kategorien untersucht

werden: Testverfahren, die Basiskompetenzen erfassen und Testverfahren, die komplexere Leseverständnisleistungen messen.

Tabelle 6: Lesetestverfahren für die Sprecher des Deutschen als Muttersprache (DaM)

Name	Abkürzung	Anwendung	Beschreibung	Autor(en)	Normierung
Salzburger Lese-Screening für die Klassenstufen 1-4	SLS 1-4	2003	Liste sehr einfacher richtiger und falscher Sätze, die möglichst schnell gelesen und nach ihrer Richtigkeit beurteilt werden sollen. Liefert das SLS 1-4 einen Wert im unteren Leistungsbereich, sollte anhand des SLRT die Problematik weiter aufgeklärt werden.	H. Mayringer H. Wimmer	Klassen-normen Reliabilität Validität N = 1867
Der Salzburger Lese- und Rechtschreibtest 1-4	SLRT	1997	Verfahren zur Diagnose von Schwächen beim Erlernen des Lesens und Schreibens; Erfassung unterschiedlicher Komponenten des Lesens: Defizite in der automatischen, direkten Worterkennung und Defizite des synthetischen, lautierenden Lesens. Erfasst werden Lesezeit und Lesefehler	K. Landerl H. Wimmer E. Moser	Alters-normen Reliabilität Validität N=2800
Würzburger Leise Leseprobe 1-4	WLLP	1998	Worterkennung, Dekodiergeschwindigkeit	P. Küspert W. Schneider	Alters-normen Reliabilität Validität N=2820
Knuspels Leseaufgaben 1-4	Knuspel-L	1998	Rekodieren und Dekodieren auf Wortebene Leseverstehen auf Satzebene	H. Marx	Alters-normen Reliabilität Validität
Diagnostischer Lesetest zur Frühd Diagnose von Lesestörungen 1-4	DLF 12	1984	Wortlesetest zur Früherfassung von Lesestörungen. Untertests umfassen Bereiche: <i>Speichern, Synthese, Analyse, Segmentierung</i>	R. Müller	Reliabilität Validität N = 320
Stolperwörter-Lesetest 1-4	STOLLE	2002	Liste mit Sätzen, in denen ein semantisch, grammatisch oder orthografisch falsches Wort eingefügt wurde. Das falsche Wort muss identifiziert werden.	W. Metze	<i>keine kontrollierte Validierung und Normierung</i>
	LUST-1	2003		H. Brügelmann A. Backhaus	Altersnormen N=7017

					Reliabilität Validität
Ein Leseverständnistest für die Klassenstufen 1-6	ELFE 1-6	2006	Leseverständnistest, der das Verständnis auf Wort-, Satz- und Textebene erfasst. Verfügbar als Computer-, Papier- und Bleistiftversion	W. Lenhard W. Schneider	Klassen- normen Reliabilität Validität N = 4893
Hamburger Lesetest für Klassenstufen 3-4	HAMLET 3-4	2006	Worterkennungstest und Leseverständnistest	R. H. Lehmann, R. Peek J. Poerschke	Alters- normen Reliabilität (ein- geschränkt) Validität (ein- geschränkt)
Züricher Lesetest	ZLT	2000	Leseflüssigkeit und -genauigkeit, qualitativ-prozessuale Analyse des Leseverhaltens	M. Linder H. Grisseman	Alters- normen N = 449
Züricher Leseverständnistest für die Klassenstufen 4-6	ZLVT 4-6	2000	Überprüfung des sinnverstehenden Lesens, orale Lesefertigkeit, Dekodieren beim stillen Lesen und handlungsmäßige Überprüfung, Darbietung in Kombination mit ZLT	H. Grisseman W. Baumberger	Alters- normen N = 300
Frankfurter Leseverständnistest für die Klassenstufen 5-6	FLVT 5-6	2008	Ein narrativer und ein expositorischer Text mit jeweils 18 Multiple-Choice-Fragen	E. Souvignier I. Trenk-Hinterberger S. Adam-Schwebe A. Goldr	Klassen- normen (jeweils 2. Halbjahr) interne Konsistenz Validität N = 2476
Salzburger Lesescreening für die Klassenstufen 5-8	SLS 5-8	2005	analog SLS 1-4	M. Auer G. Gruber H. Mayringer H. Wimmer	Klassen- normen Reliabilität Validität N = (714-850) pro Klassenstufe
Lesegeschwindigkeits- und Verständnistest für die Klassenstufen 6-12	LGV 6-12	2007	Ermittlung des Leseverständnisses und der Lesegeschwindigkeit anhand eines Lückentextes. In die Lücken müssen passende Wörter eingesetzt werden (3 Wörter stehen zur Auswahl)	W. Schneider, M. Schlagmüller M. Ennemoser	Klassen- und Schul- normen Reliabilität Validität N = 2390

Aus der Übersicht über die Lesetests für Deutsch als Muttersprache lassen sich folgende Punkte zusammenfassen:

- Die dargestellten Lesetests in den Klassen 1 bis 4 erfassen Verarbeitungsprozesse auf den unteren Prozessebenen (Dekodierungsprozesse auf der Wort- und Satzebene).

- Die Überprüfung des Leseverständnisses und der Lesegeschwindigkeit in Form von Multiple-Choice-Lesetests oder Lückentests erfolgt in den höheren Schulklassen (5/6-12). Hier werden auch höhere Prozessebenen beim Lesen überprüft.

Möchte man sowohl niedrigere als auch höhere Prozessebenen beim Lesen erfassen (vgl. dazu Kap. 2 zu den Lesemodellen), müssten einzelne Testformate kombiniert werden.

Die dargestellten Lesetests eignen sich nicht für den Einsatz im akademischen Bereich, da sie für jüngere Probanden entwickelt wurden und andere sprachliche Prozesse erfassen (basale Dekodierfähigkeit, Lesegeschwindigkeit etc.).

In Bezug auf die Gruppe der Studierenden kommen die Lesetests in Frage, die hinsichtlich Durchführung und Auswertung zeit- und kostenökonomisch sind und zugleich valide Einschätzungen der Leseleistungen der Studierenden erlauben. Im Folgenden sollen die Lesetests im Bereich Deutsch als Fremdsprache analysiert werden, da sie für erwachsene Lerner entwickelt wurden und Texte mit einer studienbezogener Thematik verwenden.

Der Lesetest, der für die Untersuchungsgruppe am ehesten in Frage kommt, ist der LGVT 6-12 (vgl. Tabelle 6). Er wurde für den Sekundarbereich konzipiert und erfasst das Leseverständnis auf der Textebene direkt und über die Lesegeschwindigkeit indirekt. Der Test gilt als Gruppenverfahren; für verschiedene Altersstufen wurden Normwerte festgelegt. Allerdings gibt es einige Gründe, die gegen die Verwendung des Tests sprechen: Der Lesetest besteht aus einem zusammenhängenden Text (ca. 4 DIN-A4-Seiten), der an mehreren Stellen vervollständigt werden muss. Es werden verschiedene Optionen zur Vervollständigung des betreffenden Satzes angeboten. Insgesamt gibt es 23 Lücken im Text mit 1.727 Wörtern. Aus dem Testmanual ist nicht ersichtlich, nach welchem Schema die Lücken in den Text eingebaut wurden. Es gibt keine Angaben bezüglich der Lückenanalyse (z.B. Lösungswahrscheinlichkeit und Reliabilität des Konstrukts). Für den Einsatz des Lesetests bei Studierenden müsste meines Erachtens nach die Thematik des Ausgangstextes geändert werden, da der Textinhalt für Studierende nicht ansprechend ist (im Text geht es um Brot und Rosenkohl). Die Auswahl geeigneter Grundlagentexte ist

zeitaufwändig und nicht einfach. Käme ein anderer Ausgangstext in Frage, müssten umfangreiche statistische Analysen vorgenommen werden, um die Gütekriterien zu sichern; außerdem hätte man für die Gruppe der Studierenden keine Vergleichs- bzw. Normwerte. Die Autoren sind der Meinung, dass die Auswahl der richtigen Antworten im Text ein gründliches, sinnverstehendes Lesen erfordere und flüchtiges Lesen ausschließe. Beim Einsatz des Tests bei Erwachsenen wäre die Tendenz denkbar, zwischen den Passagen zu springen und nur jene Sätze zu bearbeiten, die Lücken aufweisen. Studierende verfügen im Vergleich zu jüngeren Probanden über ein Repertoire an Lesestrategien, die sie beliebig einsetzen können. Außerdem wird im Text nicht gesagt, wie der Text zu lesen ist. Der Leser ist selbst für die Auswahl des Lesemodus verantwortlich.

Das Leseverständnis wird am Ende des Textes nicht überprüft; so kann auch nicht gesichert werden, ob die Probanden den Text tatsächlich verstanden haben. Da das Thema allgemein ist, können einige Lücken mithilfe des Weltwissens geschlossen werden, ohne dass der komplette Satz gelesen wird (z.B. Lücke 1, 3, 5, 16). Da nur vier Minuten zur Bearbeitung zur Verfügung stehen, kann es dazu kommen, dass der Zeitfaktor in den Vordergrund rückt und die Erfassung anderer Fähigkeiten überdeckt wird. Aus den genannten Gründen habe ich mich gegen die Verwendung des LGVT 6-12 in der vorliegenden Studie entschieden.

5.3 Überblick über Lesetests für Deutsch als Fremdsprache (DaF)

Die bisherigen Ausführungen konzentrierten sich auf die Erfassung der Lesekompetenz in der Muttersprache. Bei der Messung der Lesefertigkeiten in einer Fremdsprache spielen außer der Lesekompetenz jedoch noch weitere Teilkompetenzen eine wichtige Rolle (vgl. Lutjeharms, 2010, S. 21). „Beim Lesenlernen einer Fremdsprache muss normalerweise nicht nur die Lesefertigkeit an sich neu erworben werden, sondern eine neue Sprache.“ (Lutjeharms, 2010, S. 21). Eine neue Sprache muss nicht zwingend erlernt werden, wenn zwischen der Muttersprache und der zu erlernenden Sprache keine große Distanz besteht (z. B. wenn das gleiche Alphabet verwendet wird

und die Sprachen demselben Sprachtyp angehören). Der Erwerb einer neuen Sprache wird von bereits vorhandenen Kompetenzen in der Muttersprache beeinflusst: Erlernt etwa ein russischsprachiger Student das Deutsche als Fremdsprache, kann seine muttersprachliche Kompetenz den Erwerb der fremdsprachlichen Kompetenz durch positiven oder negativen Transfer positiv oder negativ beeinflussen. Beispiel: Obwohl es sich beim Deutschen und Russischen um Sprachen des selben Typs handelt, die strukturell große Ähnlichkeiten aufweisen, kann die Satzsyntax durch negativen Transfer zu Leseschwierigkeiten führen: Im Deutschen ist die Wortfolge weitgehend fixiert, im Russischen jedoch nicht.

Lesekompetenzen werden selten in Tests erhoben, die sich ausschließlich auf die Teilfertigkeit Lesekompetenz konzentrieren. Meist sind die Tests in Testbatterien eingefügt, in denen die vier Teilfertigkeiten getestet werden. Ich beschäftige mich im Folgenden mit den Testteilen, die sich der Erhebung der Lesekompetenz in solchen Tests widmen.

Tabelle 7: Lesetestverfahren für die Sprecher des Deutschen als Fremdsprache (DaF)

Name	Zielgruppe	Prüfungsteil Leseverstehen/ Beschreibung der Ziele	Lesetestformate Testzeit
Test Deutsch als Fremd- sprache (TestDaF)	Der TestDaF eignet sich als Sprachnachweis für 1) Studierende, die ein Studium in Deutschland aufnehmen wollen und hierfür einen Nachweis ihrer deutschen Sprachkenntnisse brauchen, 2) Studierende, die in ihrem Heimatland ihre Deutschkenntnisse nachweisen müssen, für Wissenschaftlerinnen und Wissenschaftler, die einen Aufenthalt an einer Hochschule in Deutschland planen und ihren Sprachstand überprüfen wollen,	Verstehen von Lesetexten im Hochschulalltag; Entnahme von wichtigen Informationen, gezieltes und detailliertes Verstehen von Informationen, globales Verstehen von Informationen, Verstehen von impliziten Informationen.	Der Prüfungsteil besteht aus drei Aufgaben mit 30 Items: 1 Zuordnungsaufgabe 2 Multiple-Choice-Tests (Gesamtzeit: 60 Min)

	3) für alle, die für wissenschaftliche Berufe ihre Deutschkenntnisse nachweisen müssen.		
Deutsch Sprachprüfung für den Hochschulzugang (DSH)	Mit der DSH weist der Studienbewerber nach, dass er ausreichende Kenntnisse der deutschen Sprache hat, um an einer deutschen Hochschule zu studieren.	Überprüfen von studienrelevanten Kenntnissen, zum Beispiel die Fähigkeit, in Vorlesungen Notizen zu machen, den Inhalt zu verstehen und die Informationen weiterzuverarbeiten, studienbezogene Texte zu verstehen und zu bearbeiten	<p>Aufgabenbereiche: Lesen und Grammatik</p> <p>Die Aufgaben gliedern sich in inhaltliche und grammatische Aufgaben. (Gesamtzeit: 60 Min)</p> <p>1 Lesetext mit offenen und geschlossenen Aufgaben in Form von Multiple-Choice-Fragen, Zuordnungsaufgaben, Richtig/Falsch-Aufgaben zum Text</p>
Goethe-Zertifikate	<p>Das Goethe-Zertifikat C1 befreit ausländische Studierende an verschiedenen deutschen Universitäten und Hochschulen, aber auch an vielen Studienkollegs von der sprachlichen Aufnahmeprüfung.</p> <p>Das Goethe-Zertifikat C2: (GDS) Großes Deutsches Sprachdiplom⁷⁷</p>	<p>Verstehen von einer breiten Auswahl von Lesetexten, darunter längere, komplexere Sachtexte, Kommentare und Berichte, inkl. abstrakter und komplexer Texte, Fähigkeit zum Erfassen impliziter Bedeutungen erforderlich</p> <p>Verstehen eines schriftlichen Textes hinsichtlich der darin enthaltenen Hauptaussagen, Meinungen, Einstellungen und impliziten Bedeutungen sowie Fähigkeit zur gezielten Entnahme von wichtigen Einzelheiten; Erkennen der Textstruktur; Fähigkeit, Texte schnell zu lesen, um gezielt wichtige</p>	<p>3 Leseaufgaben: Entnahme von Hauptaussagen und Einzelheiten: Aufgabe: Ergänzung von Lücken im Text (Summary Cloze): Erkennen von Meinungen und Standpunkten: Argumentativer/wertender Lesetext, Aufgabe: Eintragen richtiger Informationen in ein Raster Syntaktisch und semantisch korrekte Textergänzung: Authentischer journalistischer Text. Aufgabe: Lückentext mit 4 Multiple-Choice-Items zu jeder Frage (Gesamtzeit: 70 Min)</p> <p>4 Teile/Aufgaben Lesetexte mit Multiple-Choice-Fragen Zuordnungsaufgabe Lückentext (Gesamtzeit: 80 Min)</p>

⁷⁷ Zum 01.01.2012 hat das *Goethe-Zertifikat C2: Großes Deutsches Sprachdiplom* die drei bisherigen Prüfungen *Goethe-Zertifikat: Zentrale Oberstufenprüfung (ZOP)*, *Kleines Deutsches Sprachdiplom (KDS)* und *Großes Deutsches Sprachdiplom (GDS)* abgelöst.

		Einzelheiten zu entnehmen.	
Deutsch für Zuwanderer (DTZ)	<p>Förderung der Integration von Migrantinnen und Migranten im Sinne gesellschaftlicher Teilhabe und Chancengleichheit</p> <p>Richtet sich an Neu- und Altzuwanderer, sofern diese in der deutschen Sprache große Defizite aufweisen (nachholende Integration)</p> <p>Das Bildungs- und Qualifizierungsniveau schwankt stark: Schüler (mit Vorkenntnissen) vs. Arbeiter (ohne Vorkenntnisse, höheres Alter)</p>	<p>Überprüft wird selektierendes Lesen zur Entnahme spezifischer Informationen; globales Lesen zur Entnahme der Hauptaussagen sowie Detailverstehen; Verstehen von Informationen, aber auch Ansätze zu einer Argumentation; neben dem Verstehen von Informationen das Verstehen von Anweisungen; Wissen im Bereich Wortschatz und Strukturen sowie in Bezug auf Formen der Höflichkeit</p>	<p>5 Aufgaben:</p> <p>Multiple-Choice-Tests Richtig/Falsch-Aufgaben Zuordnungsaufgaben (Gesamtzeit: ?)</p>
Österreichisches Sprachdiplom Deutsch (ÖSD)	<p>A1-A2 Zertifikat als Nachweis von Deutschkenntnissen in Zusammenhang mit der Erteilung eines Aufenthaltstitels für Familienangehörige (Familienzusammenführung)</p> <p>B1-B2 Zertifikat als Nachweis von Deutschkenntnissen in Zusammenhang mit der Erfüllung der Integrationsvereinbarung (Daueraufenthalt) und /oder für den Erwerb der Staatsbürgerschaft⁷⁸</p> <p>C1-C2 Zertifikat als Nachweis einer sprachlichen Kompetenz, die sich bereits durch einen hohen Grad an Korrektheit und situationsspezifischer Angemessenheit auszeichnet. Besonders gut geeignet für Studienbewerber, die an einer deutschsprachigen Universität studieren möchten.</p>	<p>Überprüfen von Global- und Detailverstehen; Verstehen unterschiedlicher authentischer einfacher Texte aus Österreich, Deutschland und der Schweiz in ihrer Gesamtaussage und in ihren Einzelinhalten.</p> <p>Verstehen von Hauptinformationen, wichtigen Einzelheiten sowie Standpunkten und Meinungen. Globales, selektives und detailliertes Verstehen von unterschiedlichen Lesetexten aus Österreich, Deutschland und der Schweiz; Verstehen unterschiedlicher authentischer Texte aus Österreich, Deutschland und der Schweiz in ihrer Gesamtaussage und in ihren Einzelinhalten; Verstehen authentischer Texte zu wirtschaftsspezifischen Themen aus Österreich, Deutschland und der</p>	<p>drei geschlossene Aufgaben Ja/Nein-Aufgabe Zuordnungsaufgaben (Gesamtzeit: 25 Min)</p> <p>drei geschlossene Aufgaben 1 Zuordnungsaufgabe 2 Multiple-Choice-Tests (Gesamtzeit: 70 Min)</p> <p>4 Aufgaben Multiple-Choice-Test Zuordnungsaufgaben Lückentest (Cloze-Test) (Gesamtzeit: 90 Min)</p>

⁷⁸ Das B1-Zertifikat Deutsch ist eine Prüfung für Lernende ab 16 Jahren (Gemeinschaftsprodukt des Österreichischen Sprachdiploms Deutsch (ÖSD), des Goethe-Instituts (GI), der telc GmbH und des Lern- und Forschungszentrums der Universität Freiburg/Schweiz).

		Schweiz in ihrer Gesamtaussage und in ihren Einzelinhalten.	
--	--	-------------------------------------------------------------	--

Alle oben dargestellten Prüfungen orientieren sich an den Niveau-Beschreibungen des GER (vgl. Europarat, 2001). Für die vorliegende Studie kämen Leseaufgaben aus zwei Sprachprüfungen in Frage: aus der TestDaF- und aus der DSH-Prüfung. In beiden Prüfungen werden Multiple-Choice-Lesetests verwendet. Im Gegensatz zu entsprechenden Prüfungsteilen der DSH, die zwar auf gleichem Niveau messen, in der Regel aber nicht vorgetestet werden oder deren Testgüte in der Regel an keiner Universität überprüft wird, verfügt das TestDaF-Institut über ein differenziertes Testgütemanagement und veröffentlicht die Befunde der Vortestungen zur Überprüfung der wichtigsten Gütekriterien⁷⁹ (vgl. Arras & Grotjahn, 2002; Eckes, 2003; Arras, 2007; Kecker, 2011). Aus den genannten Gründen werden in der vorliegenden Studie die Leseaufgaben vom TestDaF genutzt. Diese werden erst an der Gruppe der Studierenden vorgetestet und, falls die entsprechenden Werte als ausreichend zu interpretieren sind, in Kombination mit den C-Tests eingesetzt.

5.4 Auswahl an Lesetests für die vorliegende Studie

5.4.1 Testkombination aus Multiple-Choice-Lesetests

Für Studierende in der vorliegenden Studie wurden Aufgaben und Testformate gewählt, die eigentlich für Fremdsprachensprecher des Deutschen entwickelt wurden, da sie zur Erfassung studienbezogener sprachlicher Fähigkeiten konzipiert wurden und universitäre Anforderungen zur Grundlage haben. Im Folgenden wird auf das Testformat Multiple-Choice eingegangen, das bei TestDaF verwendet wird. Die Tests sind so angelegt, dass gute Lerner des Deutschen als Fremdsprache auf C1-Niveau die Leseaufgaben zu 60-65 % richtig

⁷⁹ Publikationen des TestDaF-Instituts: http://www.testdaf.de/institution/tz-public_mitarbeiter.php; Publikationen über das TestDaF-Institut: http://www.testdaf.de/institution/tz-public_sonstiges.php, zuletzt gesehen am 20.06.2013.

lösen können. Bei Verwendung derselben Tests müssen wir erwarten, dass die Gruppe der muttersprachlichen Studierenden deutlich höhere Testwerte erzielt.

Begründung der Testauswahl:

Die im Weiteren beschriebenen Testaufgaben wurden für die vorliegende Studie aus folgenden Gründen gewählt: Die TestDaF-Lesetests überprüfen das Leseverständnis bei Erwachsenen konsequent mittels Multiple-Choice-Tests (MC-Tests). Die MC-Tests werden am TestDaF-Institut an einer großen Anzahl von Personen (über 1500 Probanden) erprobt und haben unterschiedliche Schwierigkeitsgrade. Die Texte für die Studierenden sollen einen mittleren Schwierigkeitsgrad⁸⁰ aufweisen, um zwischen lesestarken und leseschwachen Probanden differenzieren zu können.

Außerdem genügen die MC-Lesetests den Kriterien, die für die Verwendung im universitären Bereich an einer großen Anzahl von Studierenden von zentraler Bedeutung sind: Sie besitzen sowohl eine hohe Durchführungs- und Auswertungsobjektivität als auch eine hohe Interpretationsobjektivität. Die TestDaF-MC-Tests liefern gute Reliabilitätswerte (Cronbachs Alpha $\alpha = 0.8$ und höher)⁸¹.

Die Wahl dieses Formats – im Vergleich etwa zu offenen oder halboffenen Itemformaten, die bei Leseverstehenstests (beispielsweise bei der DSH) ebenfalls häufig eingesetzt werden – verhindert eine Überlagerung mit anderen Faktoren, welche die Untersuchungsergebnisse beeinflussen können⁸². Das bedeutet, dass durch den Einsatz geschlossener Testformate (MC-Lesetests) ausgeschlossen wird, dass die Leseleistung von der schriftlichen Leistung beeinflusst wird, was bei offenen Aufgaben nicht ausgeschlossen werden kann.

Wichtiges Merkmal dieses Aufgabentyps ist es eben, dass „die richtige Lösung nicht aktiv produziert werden muss, sondern zwischen mehreren falschen Lösungen (Distraktoren) die richtige erkannt und angekreuzt werden muss“ (vgl.

⁸⁰ Der Schwierigkeitsgrad bezieht sich auf die erbrachten Leistungen in der getesteten Gruppe der Fremdsprachenlerner (DaF).

⁸¹ Zur Bestimmung der Validität von TestDaF-Aufgaben vgl. Kecker, 2011, S. 155 ff.

⁸² Diese Erscheinung wird in der Fachliteratur Konfundierung genannt.

Bolton, 1996, S. 122; ALTE, Modul 3, S. 9 ff.). Hier muss nichts schriftlich produziert werden, was nämlich – wenn es nicht gelingt – entweder als Problem der Fähigkeit zur Schreibproduktion oder als Problem des Leseverstehens gedeutet werden kann. Das Format ist für die Auswertung besonders attraktiv, da sowohl die Durchführung als auch die Auswertung ökonomisch sind. Durch die bei der Korrektur verwendeten Schablonen oder eine maschinelle Auswertung gelten diese Lesetests als auswertungsobjektiv. Die verwendeten Schablonen stellen eine Garantie für die Möglichkeit objektiver Auswertungen dar, die natürlich nur dann valide ist, wenn die Konstruktion der Aufgaben so erfolgte, dass die Antwortmöglichkeiten eindeutig sind. Manchmal gibt es MC-Aufgaben, die auch mit einer anderen Antwort (die aber nicht zur Wahl steht) besser beantwortet werden könnten. Dies kann die Validität des Tests in Frage stellen.

Problematisch sind die MC-Leseaufgaben des TestDaF-Instituts zumindest in der Form von Ja/Nein- oder Richtig/Falsch-Antworten, weil sie zur Auflösung mit einer Chance von 50 % erraten werden können. Allerdings kann die Entwicklung dieses Testformats mit mehreren Distraktoren aufgrund der für die Konstruktion geeigneten Antworten zeitaufwendig sein. Um die Ratewahrscheinlichkeit beim Lösen zu verringern, wird im TestDaF-Institut gegenwärtig erwogen, das binäre Format aufzugeben und/oder erweiterte Antwortoptionen *Ja/Nein/ Im Text steht nichts dazu* zuzulassen.

Im Bereich Leseverstehen wird in der TestDaF-Prüfung mittels MC-Tests überprüft, „inwieweit die Studienbewerber in der Lage sind, geschriebene Texte, die im hochschulbezogenen Kontext relevant sind, angemessen zu rezipieren“ (Kecker, 2011, S. 59). Der Prüfungsteil Leseverstehen besteht aus drei Aufgaben⁸³ mit 30 Items, bei denen unterschiedliche Textsorten und unterschiedliche Itemtypen präsentiert werden. Es werden verschiedene Verarbeitungsebenen gefordert, die vom Verstehen des Gesamtzusammenhangs und bestimmter Einzelheiten bis hin zum Verstehen von impliziten Informationen reichen (vgl. TestDaF-Institut, 2009, S. 49f.).

⁸³ Das Wort *Aufgabe* wird hier in der Bedeutung *Lesetest* verwendet und wird beibehalten, weil bei der Beschreibung des Formats in der TestDaF-Anleitung das Wort *Aufgabe* figuriert.

Beschreibung von TestDaF-MC-Aufgaben

In der vorliegenden Studie werden drei Leseverstehensaufgabentypen vorgetestet, um eine geeignete Test- bzw. Textauswahl zu treffen. Im Folgenden wird die Beschreibungen dieser Aufgaben zitiert (vgl. Kecker, 2011). Die Aufgaben sind nach Schwierigkeit sortiert und werden den TestDaF-Niveaustufen⁸⁴ (TDN-Stufen) 3 bis 5 zugeordnet.

Die Aufgabe 1 weist im Vergleich zu den anderen Aufgabentypen einen geringeren Schwierigkeitsgrad auf und verlangt von den Lesenden ein allgemeines Verständnis studienbezogener Inhalte, die fachübergreifende wissenschaftliche Relevanz besitzen:

Leseverstehen 1: „Die Aufgabe 1 besteht aus einer Zuordnungsaufgabe mit 10 Items und 7 kurzen Texten (ca. 350 Wörter insgesamt), für die ca. 10 Minuten vorgesehen sind. Die Aufgabe ist der TDN-Stufe 3 zugeordnet. Die Texte bestehen aus Kurzbeschreibungen, die eine Textsorte der Alltagskommunikation repräsentieren, wie sie auch an der Hochschule häufig vorkommt, etwa in einem Vorlesungsverzeichnis, in Katalogen mit kurzen Buchbeschreibungen oder Internet-Ankündigungen. Der Lesestil besteht in selektivem oder orientierendem Lesen; die Testteilnehmer sollen nach spezifischen Informationen suchen, die durch Items vorgegeben sind, und diese den Texten entnehmen. Sprachlich und inhaltlich müssen diese Texte klar strukturiert sein, sie sollten einen eher geringen Komplexitätsgrad aufweisen und keine Fachterminologie enthalten. Dabei ist die Schwierigkeit einer solchen Aufgabe nicht zu unterschätzen, denn Kurztexte haben häufig eine hohe Informationsdichte. Darüber hinaus werden die Items so erstellt, dass sie nicht allein durch eine Titelangabe oder ein Schlüsselwort zu Beginn des Textes gelöst werden können.“ (Kecker, 2011, S. 60)

Die Aufgabe 2 wird der TDN-Stufe 4 zugeordnet und verlangt von den Testteilnehmenden Textverständnis sowohl im Gesamtzusammenhang als auch in Einzelheiten.

⁸⁴ **TDN 5** Kann geschriebene Texte aus dem studienbezogenen Alltag sowie Texte zu fächerübergreifenden wissenschaftlichen Themen, die sprachlich und inhaltlich komplex strukturiert sind, in ihrem Gesamtzusammenhang und ihren Einzelheiten verstehen und diesen Texten auch implizite Informationen entnehmen.

TDN 4 Kann geschriebene Texte aus dem studienbezogenen Alltag sowie Texte zu fächerübergreifenden wissenschaftlichen Themen, deren Struktur sich an der Allgemeinsprache orientiert, in ihrem Gesamtzusammenhang und in ihren Einzelheiten verstehen.

TDN 3 Kann geschriebene Texte aus dem studienbezogenen Alltag in ihrem Gesamtzusammenhang und in wesentlichen Einzelheiten verstehen; kann Texte zu fächerübergreifenden wissenschaftlichen Themen in Teilen verstehen. Die Informationen zu den TDN-Stufen sind auf der Internetseite des TestDaF-Instituts abrufbar: http://www.testdaf.de/teilnehmer/tn-info_nivea.php.

Leseverstehen 2: „Für diese Aufgabe wurde als Vorlage ein längerer Text (450-550 Wörter) der Textsorte Bericht ausgewählt, etwa ein Bericht aus einer Zeitung, einem Magazin o. Ä., in dem ein wissenschaftliches Thema oder ein wissenschaftlicher Zusammenhang von einem (Wissenschafts-)Journalisten für ein (Laien-)Publikum dargestellt wird. Es handelt sich hierbei nicht um wissenschaftliche Kommunikation, d. h., der Text ist kein wissenschaftlicher Text. Der Bezug zum Kommunikationsbereich Hochschule wird über das Thema etabliert, nicht über die Textsorte. Der Text soll einen klaren Aufbau haben, komplexe sprachliche Strukturen enthalten, jedoch wenig Fachterminologie. Fachtermini, die vorkommen, müssen kenntlich gemacht [sein] im Text selbst oder in einer Fußnote explizit erklärt sein. Der Text soll über ein deutlich höheres Maß an Informationsdichte und abstrakter Information verfügen als die Texte in LV1, aber für einen Laien verständlich sein. Zu dem Text gehören 10 Multiple-Choice-Items mit drei Antwortoptionen. Für die Lösung der Items sind 20 Minuten vorgesehen. Die Items zielen vorwiegend auf das Erkennen von Gesamtzusammenhängen, Hauptaussagen und Einzelheiten ab (globales und detailliertes Verstehen), dem Text sind jedoch vereinzelt auch implizite Bedeutungen zu entnehmen.“ (Kecker, 2011, S. 60)

Die Aufgabe 3 ist die anspruchsvollste von allen beschriebenen Leseaufgaben. Sie ist sprachlich und inhaltlich komplex strukturiert und verlangt von den Lesenden ein globales Textverständnis sowohl auf Textebene als auch eine implizite Informationsentnahme.

Leseverstehen 3: „Die dritte Aufgabe im Prüfungsteil Leseverstehen ist auf die TDN-Stufe 5 ausgerichtet. Auch in dieser Aufgabe muss ein längerer Text bearbeitet werden (550-650 Wörter). Der Text ist ein wissensvermittelnder Text. Dabei handelt es sich um Textsorten aus dem Kommunikationsbereich Hochschule, die der Vermittlung (und Erarbeitung) wissenschaftlicher Inhalte dienen, z. B. Texte aus Hochschullehrbüchern und Einführungen, aber auch Texte aus Hochschulmagazinen, in denen z. B. neue wissenschaftliche Erkenntnisse dargestellt werden. Kennzeichnend für diese Texte ist u. a., dass sie nicht der (fachinternen) Kommunikation zwischen Experten dienen, sondern der Kommunikation zwischen einem (Fach-)Experten und jemandem, der in dem angesprochenen Fachgebiet einen geringeren Grad an Spezialisierung aufweist, also einem Studenten oder einem Akademiker aus einem anderen Fachgebiet. Der Text muss einen klaren Aufbau haben und durch wissenschaftliche textuelle Strukturen geprägt sein (z. B. wissenschaftssprachliche Terminologie und Wortbildung, argumentative Darlegungen, Nominalisierungen, komplexe Satzstrukturen). Der Anteil an Fachsprache sollte der Textsorte entsprechend begrenzt sein. Fachtermini müssen entweder im Text definiert werden oder aus dem Kontext erschlossen werden können. Der Text soll über eine größere Informationsdichte, ein höheres Maß an abstrakter Information sowie über komplexere syntaktische Strukturen verfügen als die Textvorlage für den LV2. Zu dem Text gehören 10 Items mit den Antwortoptionen Ja/Nein/Text sagt dazu nichts. Für die Lösung sind 20 Minuten vorgesehen. Die Items beinhalten einen erhöhten Schwierigkeitsgrad, da die Abgrenzung zwischen Optionen 'Nein' und 'Text

sagt dazu nichts' schwerfällt und Lösungen im Text dazu häufig impliziten Informationen zu entnehmen sind.“ (Kecker, 2011, S. 60f.)

Für die Vorerprobung der oben beschriebenen Aufgaben mit den Adressaten meiner Studie wurden drei Modellsätze⁸⁵ ausgewählt, die einen Schwierigkeitsgrad zwischen $p = 0.621$ und 0.675 ⁸⁶ aufweisen. Diese Werte sind für die Gruppe der Fremdsprachenlerner als mittelschwer bzw. leicht einzuschätzen. Die dargestellten Werte zeigen, dass mehr als 60 % der Studienbewerber die MC-Aufgaben korrekt lösten. In der Gruppe der Studierenden werden, wie oben bereits erwähnt, höhere Werte erwartet. Allerdings wurde die Aufgabe durch Verkürzung der Bearbeitungszeit erschwert.

Erprobung von MC-Lesetests an Studierenden

Zur Erprobung in der Gruppe der Studierenden wurden vier verschiedene MC-Lesetests gewählt: **Lesetext A** *Wann arbeiten wir am besten?* (TestDaF T005, Lesetest 02); **Lesetext B** *Lebenslanges Lernen* (TestDaF T005, Lesetest 03); **Lesetext C** *Schlafen und Träumen* (TestDaF T009, Lesetest 03) und **Lesetext D** *Lust oder Last?* (TestDaF Modellsatz 03; Lesetest 02) (vgl. Anhang A, Teil 2.4).

Die gewählten Texte haben die gleiche Länge und unterscheiden sich nur in Bezug auf ihre Thematik. Die Leseaufgabe besteht darin, nach dem Lesen von den angebotenen Antwortoptionen (*Ja/Nein/ Im Text steht nichts dazu*) korrekte Antworten anzukreuzen. Die Bearbeitungszeit für Studierende wurde statt 20 Minuten (bei den Fremdsprachenlernern) auf 10 Minuten verkürzt, um die Leseanforderung zu erhöhen. In Kombination mit den beschriebenen MC-Tests wurde eine MC-Zuordnungsaufgabe dargeboten, die eine Zuordnung von bestimmten Informationen verlangt.

⁸⁵ Als Modellsatz wird der Prüfungsteil zum Bereich Leseverstehen genannt, der aus drei Leseaufgaben besteht.

⁸⁶ Die Schwierigkeiten wurden im Rahmen von TestDaF-Untersuchungen gerechnet und wurden von den Mitarbeitern des TestDaF-Instituts zur Verfügung gestellt. In Absprache wurden für die Testungszwecke folgende Modellsätze verwendet: T005, T009 und 03. Der Modellsatz 03 kann auf der Seite des TestDaF-Instituts heruntergeladen werden: http://www.testdaf.de/teilnehmer/tn-vorbereitung_modellsatz3.php, die beiden anderen Testhefte (T005 und T009) hat das TestDaF-Institut freundlicherweise zur Unterstützung meiner Untersuchung zur Verfügung gestellt. Texte mit einem höheren Schwierigkeitsgrad oder Texte, die wegen zu hohem Schwierigkeitsgrad aus den TestDaF-Testungen aussortiert wurden und somit eine bessere Tauglichkeit für Studierende hätten, gab es keine.

Diese Aufgabe wird im TestDaF-Prüfungsteil Lesen als leichte Aufgabe klassifiziert und der TDN-Stufe 3 zugeordnet (vgl. Beschreibungen der TestDaF-Aufgaben bei Kecker, 2011, S. 60f.). An dieser Stelle wird entschieden, die Aufgaben mit verschiedenen Schwierigkeitsniveaus in die Testbatterie aufzunehmen. Auf diese Weise hat der Test eine sogenannte gestaffelte Schwierigkeit. Für die Bearbeitung der MC-Zuordnungsaufgabe wurden den Studierenden statt 10 Minuten (Fremdsprachenlerner) 5 Minuten zur Verfügung gestellt. So wurde die Testaufgabe, die von allen die leichteste ist, dem Leistungsniveau angepasst.

Ergebnisse der Vortestungen zur Auswahl der Lesetests:

Der Testteil *Leseverstehen* in der vorliegenden Studie besteht aus 20 Items, d.h. der MC-Lesetest, der das Lesen eines längeren Textes vorsieht, besteht aus 10 Fragen zum Text, ebenso wie die Zuordnungsaufgabe, die 10 Fragen umfasst. Für die Gruppe der Studierenden hat sich die Kombination aus einem MC-Lesetest (LA03) und einer MC-Zuordnungsaufgabe (LA01) aus dem Modellsatz T009 als geeignet erwiesen (vgl. Berechnungen im Anhang A, Teil 2.4). Andere für die Testung verwendete MC-Lesetests erreichten in der getesteten Gruppe keine ausreichende Reliabilität in Kombination mit der MC-Zuordnungsaufgabe und wurden daher aus der Testung eliminiert (vgl. Berechnungen im Anhang A, Teil 2.7). Die berechneten Reliabilitäten für die Lesetestkombination A beträgt Cronbachs Alpha $\alpha = 0.794$; Lesetestkombination B den Wert $\alpha = 0.779$; Lesetestkombination C den Wert $\alpha = 0.809$ und die Lesetestkombination D den Alpha-Werte von $\alpha = 0.685$. Aufgrund einer höheren Reliabilität (Cronbachs Alpha $\alpha = 0.809$) im Vergleich zu den anderen vorgetesteten Lesetestkombinationen wurde die Lesetestkombination C gewählt, die aus einem MC-Lesetest mit dem Titel *Schlafen und Träumen* und einer MC-Zuordnungsaufgabe mit dem Titel *Infos für die Studien- und Berufsplanung* bestand (vgl. Anhang A, Teil 2.4). Die ausgewählten Lesetests wurden anschließend zusammen mit vier C-Tests (4 C-Tests, 1 MC-Lesetest, 1 MC-Zuordnungsaufgabe) im Rahmen mehrerer Veranstaltungen an der Universität Duisburg-Essen 171 Studierenden vorgelegt. Der Durchführungszeitraum der Testung betrug 60 Minuten (zu den Ergebnissen vgl. Kap. 7). Die Reliabilität,

gemessen in der Gruppe der 171 Studierenden, erreichte für die Lesetestkombination C einen Alpha-Wert von $\alpha = 0.872$, was als sehr zufriedenstellend zu interpretieren ist (vgl. Anhang A, Tabelle 130).

In Tabelle 7 werden die beschriebenen Tests in Bezug auf die in der Gruppe der Studierenden ermittelten Schwierigkeitsgrade dargestellt.

Tabelle 8: Testkombination aus C-Tests und Multiple-Choice-Lesetests (N=171)

Name des Testformats	Thema	Schwierigkeit
4 C-Tests (100 Items) (Sprachkompetenzprojekt Diagnose & Förderung)	Pädagogik/Psychologie	(T1) $p = 0.61$ (T2) $p = 0.55$ (T3) $p = 0.54$ (T4) $p = 0.50$ Gesamt $p = 0.62$
1 MC-Lesetest (10 Items) 1 MC-Zuordnungsaufgabe (10 Items) TestDaF-Prüfung	allg. Psychologie Studiumsplanung	$p = 0.68$ (10 Items) $p = 0.75$ (10 Items) Gesamt $p = 0.71$
Gesamtestbatterie (6 Tests)		$p = 0.77$

Aus der Tabelle ist ersichtlich, dass die verwendeten Tests (C-Tests und MC-Lesetests) für die Studierenden eher als leichte Tests zu interpretieren sind. Betrachtet man die Testschwierigkeiten für C-Tests und MC-Lesetests separat, so sind die C-Tests schwieriger zu lösen als die MC-Leseaufgaben. Die Zuordnungsaufgabe ist, wie erwartet, die leichteste in der Testbatterie und wird von 70 % der Studierenden korrekt und vollständig gelöst.

Analysiert man die verwendeten Tests in Bezug auf ihre Aufgabenstellung und die getesteten Teilkompetenzen, so ergibt sich folgendes Bild:

Tabelle 9: Analyse der zu messenden sprachlichen Teilkompetenzen mit verschiedenen Testverfahren

C-Tests	MC-Lesetest	MC-Zuordnungsaufgabe
Verarbeitung auf der Wort-, Satz- und (eingeschränkt) auf der Textebene ⁸⁷	Decodierungsfähigkeit auf der Wort-, Satz- und Textebene	Decodierungsfähigkeit auf der Wort-, Satz- und z.T. auch auf der Textebene
Aktivierung des sprachlichen Wissens aus den Bereichen Lexik, Semantik, Grammatik, Syntax	Schnelle selektive Lokalisierung von Informationen in Abhängigkeit vom Leseziel	Selektives oder orientierendes Lesen (Suche nach spezifischen Informationen, die durch Items vorgegeben sind)
Aktivierung von den mit dem Lesen verbundenen hierarchieniedrigen (Wortidentifikation, lokale Kohärenzherstellung) und hierarchiehöheren Prozessen (z.B. globale Kohärenzbildung, Erkennen von Superstrukturen, Vorwissen, Arbeitsgedächtnis)	Verstehen der Detailinformationen und Hauptaussagen (globales und selektives Lesen)	Erkennen von sprachlichen und inhaltlichen Strukturen im Text (lokale und globale Kohärenz)
Erkennen von einfachen und komplexen Textbezügen (wort-, satz- und textübergreifend)	Verbindung von Einzelaussagen und Argumentation zu einem Gesamtverständnis von Textpassagen (z. B. Ursache und Wirkung erkennen, Gegensätze und Ähnlichkeiten identifizieren)	Verwendung von den mit dem Lesen verbundenen hierarchieniedrigen Prozesse
Lese- und Verarbeitungsgeschwindigkeit (C-Tests sind zeitlich limitiert), Lesegenauigkeit (inhaltlich richtig vervollständigte Lücken werden mit einem Punkt gewertet)	Verwendung von den mit dem Lesen verbundenen hierarchieniedrigen (Wortidentifikation, lokale Kohärenzherstellung) und hierarchiehöheren Prozessen (globale Kohärenzbildung, Erkennen von Superstrukturen, Vorwissen, Arbeitsgedächtnis)	
Globales und detailliertes Lesen (Top-Down- und Bottom-Up-Prozesse bei der Bearbeitung der Lücken werden aktiviert)		

⁸⁷ Die Einschränkung auf der Textebene betrifft das C-Testformat. Wegen der relativ kurzen Textabschnitte, aus denen die C-Tests bestehen, ist es eingeschränkt möglich, die Leseprozesse auf den höheren Prozessebenen im Sinne der Lesemodelle von Richter & Christmann und Rosebrock & Nix zu erfassen (vgl. dazu Kap. 2.2.2 und Kap. 2.2.3).

Die oben genannten Punkte zu den einzelnen Tests zeigen, dass C-Tests und Lesetests Überschneidungspunkte, aber auch Abweichungen in Bezug auf das Testkonzept aufweisen. MC-Lesetests dienen vor allem der Überprüfung des Leseverständnisses, C-Tests dagegen fokussieren sich stärker auf andere sprachliche Bereiche wie z. B. Lexik, Semantik oder Grammatik. Grammatik spielt eine untergeordnete Rolle, trägt aber zur Rekonstruktion des Sinnes auf der Satzebene bei. Stärker wird Grammatik bei C-Tests mit Wortendtilgungen erfasst. Für eine erfolgreiche Rekonstruktion von C-Tests ist der Einsatz verschiedener Lesestrategien notwendig. Das Leseverständnis und die Kontrolle über den Einsatz bestimmter Lesestrategien werden mit diesem Format sekundär mitgetestet. Die C-Tests bestehen aus relativ kurzen Texten; daher können Superstrukturen im Text oder globale Kohärenzherstellung nur eingeschränkt überprüft werden.

Den MC-Lesetests liegen in der Regel wesentlich längere Texte zugrunde (ca. 450-600 Wörter) als den C-Tests. Das hat sicher Auswirkungen auf den Leseprozess und belastet die Kapazität des Arbeitsgedächtnisses sehr unterschiedlich. Ein wesentlicher Unterschied zu den MC-Lesetests ist darin zu sehen, dass die Aufgabenstellung beim C-Test gar nicht darauf abzielt, den Text als Quelle der Information zu nutzen, sondern dass im Text enthaltene Informationen vervollständigt werden müssen. Der Leseprozess wird bei MC-Tests stark von der jeweiligen Aufgabenstellung bestimmt. Werden die Fragen zum Text dem Leser vorab vorgelegt, so erfolgt das Lesen stärker selektiv, als wenn man die Fragen erst nach dem Lesen beantworten lässt.

In der vorliegenden Untersuchung wurden die Fragen gestellt, *nachdem* der Text gelesen worden war. Wegen seiner Länge und der behandelten Thematik müssen die Probanden bei der Rezeption der ausgewählten MC-Lesetests z. B. komplexere Strukturen identifizieren, Gegensätze oder Widersprüche aufdecken und Argumentationen erkennen.

Nach den beschriebenen Bereichen besitzt die MC-Zuordnungsaufgabe mehr Überschneidungspunkte mit den C-Tests als die MC-Lesetests. Die MC-Zuordnungsaufgabe fokussiert ebenso wie der C-Test stärker die

Decodierungsfähigkeit auf der Wort- und Satzebene, da die Textabschnitte i.d.R. sehr kurz sind. In beiden Tests verwenden die Probanden Techniken des selektiven Lesens, weil sie nach spezifischen Informationen suchen. Bei der MC-Zuordnungsaufgabe sind es Textabschnitte, die mit Hilfe der Techniken des detaillierten und suchenden Lesens richtigen Aussagen zugeordnet werden. Ähnliche Strategien scheinen auch bei der Rekonstruktion der C-Tests aktiviert zu werden: Die Probanden suchen nach geeigneten Wörtern, die unter Einbeziehung des Satz- oder Textkontextes in die jeweiligen Lücken eingesetzt werden können. Die in der Lücke vorgegebenen Wortteile und die im Satz vollständig stehenden Wörter sollen im Kontext dabei helfen, den Satzsinn zu rekonstruieren. Sowohl in den C-Tests als auch in der MC-Zuordnungsaufgabe werden in starkem Maße hierarchieniedrige Prozesse verwendet. Allerdings kommen auch hierarchiehöhere Prozesse eingeschränkt zum Einsatz, wie ich an anderer Stelle gezeigt habe (vgl. Kap. 6).

In Kap. 7 werden Hypothesen in Bezug auf die Korrelationen zwischen den vorgestellten Tests formuliert und überprüft; daher wird an dieser Stelle auf diesen Bereich nicht weiter eingegangen (vgl. auch die Grunddaten und statistischen Berechnungen in Anhang A, Teil 3).

5.4.2 Testkombination aus Stolperwörter-Lesetests

Die oben genannten Kriterien der Praktikabilität und Ökonomie scheint auch der Stolperwörter-Lesetest, im Folgenden Stolperwörtertest genannt, zu erfüllen. Er wurde als Testformat von Wilfried Metze (2002) zur Erfassung von Lesefertigkeit und -fähigkeit von Schulkindern im deutschen Sprachraum konzipiert. Der Lesetest erfüllt zwei Kriterien, die für die Praxis sehr wichtig sind: Zum einen ist er sowohl in der Durchführung⁸⁸ als auch in der Auswertung⁸⁹ zeitökonomisch, und zum anderen liefert er in Bezug auf das Leseverständnis valide Werte⁹⁰.

⁸⁸ Die durchschnittliche Durchführungszeit eines Stolperwörtertests inkl. Instruktion beträgt ca. 20 Minuten.

⁸⁹ Die Auswertung erfolgt über Schablonen. Es werden Rohwerte für die bearbeiteten Sätze ermittelt. Die Auswertung erhöht die Objektivität des Testverfahrens.

Das Testformat fand große Verbreitung an Grundschulen; entsprechende Daten wurden in der Studie STOLLE (*Stolperwörter-Lesetests*)⁹¹ in Berlin erhoben (Metze, 2005)⁹². 2003 wurde der Stolperwörtertest im Rahmen der Studie LUST (*Leseuntersuchung mit dem Stolperwörter-Test*) an der Siegener Universität in nordrhein-westfälischen Grundschulklassen eingesetzt⁹³. Auf diese Weise entstand eine zweite Studie, die sich auch der Erprobung der Stolperwörtertests widmete. Der Hauptunterschied zwischen den Studien besteht darin, dass sich LUST auf die Erhebung zur Jahresmitte und STOLLE auf das Schuljahresende bezieht. Beide Studien erbrachten konkurrierende Ergebnisse, da die durchschnittlichen Leseleistungen, die bei LUST erhoben wurden, bereits zum Schulhalbjahr höher waren als die Ergebnisse aus der STOLLE-Studie ein halbes Jahr später. Die Ergebnisse deuten darauf hin, dass es bei den Tests auch noch andere Unterschiede gab als nur den Erhebungszeitpunkt. Diesen Unterschied konnten die Autoren nicht erklären und vermuteten, dass er von der Stichprobenzusammensetzung beeinflusst wurde. Den Testverwendern wird empfohlen, sich für einen festen Erprobungszeitpunkt (Schuljahresmitte oder Schuljahresende) zu entscheiden, da unterschiedliche Normwerte für die gleiche Klassenstufe vorliegen und man sich an *einem* Wert orientieren sollte (vgl. Backhaus et al., 2003, S. 3).

Zum Inhalt: Der Stolperwörtertest sieht vor, dass in einer Reihe von Sätzen ein Wort aus dem Satzkontext als ‚Störer‘ identifiziert wird.

Beispiel: *Heutige Schulklassen bestehen aus SchülerInnen und Schülern ohne mit Migrationshintergrund.*

Lösung: *Heutige Schulklassen bestehen aus SchülerInnen und Schülern ~~ohne~~ mit Migrationshintergrund.*

Da die Identifikation dieser Störer nur auf der Basis eines sinnentnehmenden Lesens möglich ist, erfasst das Testverfahren die Lesefähigkeit der getesteten Personen. Aufgrund seiner Testkonzeption ermöglicht der Stolperwörtertest

⁹⁰ Es wurden Korrelationen mit anderen Verfahren zur Erfassung der Leseleistung mit weiteren Leseverständnistests (Worttest O40 $r=.41$ bis $.76$; Hamlet 3/4 Leseverständnis $r=.61$ bis $.87$) festgestellt.

⁹¹ <http://www.wilfriedmetze.de/html/stolper.html>, zuletzt gesehen am 05.04.2013.

⁹² http://www.wilfriedmetze.de/Stolle_2005.pdf, zuletzt gesehen am 05.04.2013.

⁹³ Insgesamt wurden im Nordrhein-Westfalen über 20.864 Kinder getestet (vgl. Brügelmann, 2003, S. 6).

Einblicke in folgende Bereiche des Lesens: Lesegenauigkeit⁹⁴, Lesegeschwindigkeit⁹⁵, Decodierungsfähigkeit und Leseverständnis auf der Wort- und Satzebene. Das Leseverständnis auf der Textebene wird aufgrund des Formats⁹⁶ nicht miterfasst, weil einzelne Sätze miteinander nicht zusammenhängen. Der Stolperwörtertest wurde nicht nur an Schulkindern erprobt, sondern auch an Erwachsenen (vgl. Brügelmann, 2004). Eine der getesteten Gruppen bildeten Lehrer und Lehramtsstudierende. In der Studie zeigte sich, dass einige Lehrer so lese schwach waren, dass sie teilweise keine besseren Leseleistungen erzielten als einige Grundschüler. Getestet wurden in der Studie 75 Meisterschüler aus Kursen einer Handwerkskammer, 55 Lehramtsstudierende und 181 Lehrer. Die Studie erbrachte folgenden zentralen Befund: Das Mindestniveau der Handwerkergruppe von 3 richtigen Sätzen pro Minute erreichen bereits 68 % der Zweitklässler und 99 % der Viertklässler (jeweils zur Mitte des Schuljahres). Betrachtet man die folgende Tabelle, so ist der Leistungsanstieg gut zu erkennen. Von Lehramtsstudierenden werden durchschnittlich 17 bis 18 Sätze pro Minute richtig gelesen und von Lehrern 18 bis 19. Vergleicht man jedoch die Leistungen der leistungsschwächeren Lehramtsstudierenden und Lehrer, können ihre Testwerte mit denen lese starker Dritt - und Viertklässler verglichen werden.

Tabelle 10: Einsatz des Stolperwörtertests für Erwachsene

Tab. 1 ⁵ : Leseleistungen verschiedener Gruppen (Stand: 04-02-03)					
	Mitte ⁶ 2. Klasse N = 6.654	Mitte ⁷ 4. Klasse N = 6.415	Hand- werker N = 75	Lehramts- Studierende N = 55	Lehrer- Innen N = 180
Richtige Sätze pro Minute Durchschnitt	4	8	11	17-18	18-19

Quelle: Brügelmann, 2004, S. 2.

⁹⁴ Die summierte Zahl der fehlerhaft bearbeiteten Sätze im Test gibt Auskunft über die Genauigkeit der gelesenen Sätze.

⁹⁵ Die Bearbeitung des Tests findet unter Zeitdruck statt, sodass Informationen zum Lesetempo gewonnen werden (Anzahl der bearbeiteten bzw. richtigen Sätze).

⁹⁶ Der Test besteht aus mehreren Sätzen, die kontextuell miteinander nicht verbunden sind; daher ist eine Erfassung des Leseverständnisses aus der Textebene nicht möglich.

Die Befunde der dargestellten Untersuchung zeigten, dass der schulische Stolperwörtertest auch an der Gruppe der Erwachsenen getestet werden kann. In der Studie von Brügelmann ist die Anzahl der getesteten Lehramtsstudierenden sehr klein und die Interpretation der Ergebnisse in Bezug auf die Leistungen der Grundschüler und Lehramtsstudierenden bzw. Lehrer problematisch. Da der Test jedoch den wichtigsten Gütekriterien entspricht, habe ich mich dazu entschieden, in der vorliegenden Studie den schulischen Stolperwörtertest auch an der Gruppe der Studierenden zu erproben.

In der LUST-Studie werden mehrere Testvarianten kostenfrei zur Auswahl angeboten. Für Testzwecke im Rahmen dieser Untersuchung wurde die **Version A⁹⁷** gewählt (*Originalfassung von Metze (2000) für Klassen 2 bis 4*). Dieser Stolperwörtertest wird im Folgenden zur besseren Orientierung *schulischer Stolperwörtertest* genannt. Der Test besteht aus 60 Items, die innerhalb von 5 Minuten gelöst werden müssen (vgl. die Aufgabenstellung).

Einsatz des Stolperwörtertests an Studierenden:

Getestet wurde der Stolperwörtertest in einer Einführungsveranstaltung. An der Testung nahmen 223 Studierende teil. Wie erwartet, war der Stolperwörtertest für die Untersuchungsgruppe zu einfach. Von 60 vorgelegten Items wurden 10 von allen Probanden zu 100 % gelöst. Diese Items enthalten keine Aussagekraft für die Gesamtgruppe. Der Stolperwörtertest erreicht einen sehr geringen Reliabilitätswert von $\alpha = 0.471$, was für diagnostische Zwecke als nicht ausreichend zu interpretieren ist. Daher hielt ich es für angebracht, den Stolperwörtertest an das Niveau der Studierenden anzupassen, da ich ihn aufgrund seiner erwiesenen Gütekriterien grundsätzlich auch für die in meiner Studie zu testenden Probanden für geeignet halte. Folgende Punkte sollten bei einer Anpassung beachtet werden. Der modifizierte Stolperwörtertest soll:

- a) Die Grundprinzipien des Testformats übernehmen,
- a) aus weniger Items bestehen, als dies von Metze und Brügelmann vorgeschlagen wird,⁹⁸

⁹⁷ <http://www2.agprim.uni-siegen.de/lust/Anlage%204.pdf>

⁹⁸ Die Ergebnisse der statistischen Analysen zeigen, dass für das Erreichen hoher Reliabilitätswerte 35 Items genügen.

- b) inhaltlich anspruchsvollere Sätze aufweisen,
- c) ein für die Studierenden relevantes Thema behandeln,
- d) komplexe sprachliche Strukturen enthalten (z. B. Wortbildungen, syntaktische Strukturen).

Modifizierter Stolperwörtertest:

Der modifizierte Stolperwörtertest enthält statt 60 nur 35 Items, ist aufgrund der hochschulbezogenen Thematik inhaltlich anspruchsvoller und weist auf Wort- und Satzebene komplexere sprachliche Strukturen auf. Die im Stolperwörtertest verwendeten Sätze stammen aus den Fachzeitschriften *Pädagogik* und *Psychologie heute*, die auch als Grundlagenquellen für die Entwicklung der C-Tests für Lehramtsstudierende dienen.

In einer kleinen Gruppe wurde zunächst überprüft, wie viel Zeit benötigt wird, um den Tests komplett zu lösen. Die Voruntersuchung zeigte, dass der Test in 5 Minuten gelöst werden kann. Aufgrund dieser Ergebnisse wurde entschieden, den Studierenden 5 Minuten zur Verfügung zu stellen (zur Entwicklung des Stolperwörtertests vgl. Anhang A, Teil 2.3).

Mit dem Einsatz des Stolperwörtertests soll überprüft werden, wie schnell die Studierenden lesen und wie viele Fehler sie identifizieren können.

Einsatz des modifizierten Stolperwörterlesetests:

Der modifizierte Stolperwörtertest wurde im Rahmen einer Einführungsveranstaltung insgesamt 397 Studierenden vorgelegt. Wie auch die MC-Lesetests wurde er zusammen in einem Testset mit vier C-Tests dargeboten. Getestet wurde in zwei unabhängigen Gruppen; die Aufteilung war relevant für die Reihenfolge der Bearbeitung von Stolperwörtertests. Eine Gruppe musste den Stolperwörtertest vor den C-Tests bearbeiten (N = 246) und die andere nach den C-Tests (N = 151). In beiden Gruppen ist die berechnete Reliabilität zufriedenstellend und beträgt $\alpha = 0.93$ (N=246) und $\alpha = 0.76$ (N = 151). Auffällig ist, dass die Reliabilität sinkt, wenn die Reihenfolge der Bearbeitung verändert wird (vgl. dazu Kap. 7), d. h., wenn der C-Test als erster Test bearbeitet werden muss. Im Folgenden wird auf die Schwierigkeiten der jeweiligen Tests

eingegangen (zur den entsprechenden Daten und Berechnungen vgl. Anhang A, Teil 3).

Tabelle 11: Testset aus C-Tests zuerst und Stolperwörtertest danach (N=151)

Name des Testformats	Thema	Schwierigkeit
4 C-Tests (100 Items) (Sprachkompetenzprojekt Diagnose & Förderung)	Pädagogik/Psychologie	(T1) $p = 0.63$ (T2) $p = 0.68$ (T3) $p = 0.65$ (T4) $p = 0.60$ Gesamt $p = 0.64$
Modifizierter Stolperwörtertest (35 Items)	Pädagogik/Psychologie	$p = 0.68$
Gesamtbatterie (135 Items)	Pädagogik/Psychologie	$p = 0.69$

Auffallend sind die unterschiedlichen Testschwierigkeiten in den Tabellen 11 und 12, was bedeutet, dass in den getesteten Gruppen der Anteil der leistungsstarken und leistungsschwachen Studierenden unterschiedlich groß ist. In der ersten Gruppe bereiten sowohl der C-Test als auch der Stolperwörtertest keine gravierenden Schwierigkeiten bei der Bearbeitung. Beide Tests werden von 65 % der Studierenden korrekt gelöst. In der zweiten Gruppe scheint der C-Test (besonders der erste Teilstest) für die Gruppe schwierig zu sein. Der Stolperwörtertest wird jedoch in der zweiten Gruppe von 60 % der Studierenden richtig gelöst (vgl. ausführliche Beschreibungen dazu in Kap. 7).

Tabelle 12: Testset aus Stolperwörtertest zuerst und C-Tests danach (N=246)

Name des Testformats	Thema	Schwierigkeit
Modifizierter Stolperwörtertest (35 Items)	Pädagogik/Psychologie	$p = 0.61$
4 C-Tests (100 Items) (Sprachkompetenzprojekt Diagnose & Förderung)	Pädagogik/Psychologie	(T1) $p = 0.44$ (T2) $p = 0.55$ (T3) $p = 0.53$ (T4) $p = 0.52$ Gesamt $p = 0.51$

Gesamtbatterie (135 Items)	Pädagogik/Psychologie	p = 0.54
-------------------------------	-----------------------	----------

Beschreibt man einzelne Tests in Bezug auf die zu testenden Teilkompetenzen, so ergeben sich folgende Punkte, die sich z. T. überschneiden.

Tabelle 13: Mit den C-Tests und dem Stolperwörtertest getestete sprachliche Teilkompetenzen

C-Tests	Stolperwörtertest für Erwachsene
Verarbeitung auf der Wort-, Satz- und (eingeschränkt) auf der Textebene ⁹⁹	Decodierungsfähigkeit auf der Wort- und Satzebene (zum Beispiel phonologisch, lexikalisch, syntaktisch, semantisch)
Aktivierung des sprachlichen Wissens aus den Bereichen Lexik, Semantik, Grammatik, Syntax	Identifikation von eingebauten Fehlern aus folgenden Bereichen: Inhalt, Semantik, Orthografie, Syntax, Grammatik
Aktivierung von den mit dem Lesen verbundenen hierarchieniedrigen (Wortidentifikation, lokale Kohärenzherstellung) und hierarchiehöheren Prozessen (globale Kohärenzbildung, Erkennen von Superstrukturen, Vorwissen, Arbeitsgedächtnis)	Lese- und Verarbeitungsgeschwindigkeit (der Stolperwörtertest ist zeitlimitiert)
Erkennen von einfachen und komplexen Textbezügen (wort-, satz- und textübergreifend)	Lesegenauigkeit (richtig erkannte eingebaute 'Störer' werden mit einem Punkt gewertet)
Lese- und Verarbeitungsgeschwindigkeit (C-Tests sind zeitlich limitiert), Lesegenauigkeit (inhaltlich richtig vervollständigte Lücken werden mit einem Punkt gewertet)	
Globales und detailliertes Lesen (Top-Down- und Bottom-Up-Prozesse bei der Bearbeitung der Lücken werden aktiviert)	

Wie aus den oben genannten Punkten ersichtlich, überschneiden beide Tests sich in vier wesentlichen Punkten: *Decodierungsfähigkeit auf der Wort- und Satzebene*, *Aktivierung des sprachlichen Wissens* im Bereich Syntax, Grammatik und Semantik, *Lesegenauigkeit* (Richtigkeit) und *Lesegeschwindigkeit*, da beide

⁹⁹ Die Einschränkung auf der Textebene betrifft das C-Testformat. Wegen der relativ kurzen Textabschnitte, aus denen die C-Tests bestehen, ist es eingeschränkt möglich, die Leseprozesse auf den höheren Prozessebenen im Sinne von Lesemodellen von Richter & Christmann und Rosebrock & Nix zu erfassen (vgl. dazu Kap. 2.2.2 und Kap. 2.2.3).

Tests unter Zeitdruck gelöst werden müssen. Der Punkt *Lesegenauigkeit* bedeutet in Bezug auf die C-Tests, dass die Lesenden die Lücken grammatisch und inhaltlich korrekt rekonstruieren. In Bezug auf den Stolperwörtertest bedeutet es eine Identifikation der im Satz eingebauten Fehler, die unterschiedlichen Bereichen (z. B. nur dem Bereich Grammatik oder nur dem Bereich Syntax) zugeordnet werden. Beim C-Test ist der Leser aktiv; er sucht selbst nach einem geeigneten Wort. Beim Stolperwörtertest dagegen muss er aus den vorgegebenen Varianten wählen – auch das verlangt produktive Fähigkeiten, die sich allerdings nicht auf der Schreibebebene konkretisieren.

Aus den oben genannten Gründen wird angenommen, dass bei dem Vergleich beider Testformate statistisch signifikante Korrelationen erreicht werden können. Ich nehme jedoch an, dass die Korrelationen sich nur in einem mittleren Bereich bewegen, da der Stolperwörtertest aufgrund seines Testformats Lesestrategien nicht erfasst, die mit höheren Prozessebenen verbunden sind. Der Stolperwörtertest testet die Lesekompetenz nur auf Wort- und Satzebene. Der C-Test dagegen geht über diese Grenzen hinaus und fordert auch die Aktivierung hierarchiehöherer Prozessen, die nur auf der Textebene möglich sind. Beim Stolperwörtertest sind einzelne Sätze im Test inhaltlich nicht miteinander verbunden; dadurch ist, wie bereits in vorherigen Kapiteln erwähnt, die Überprüfung des Gesamtkontextes und anderer mit dem Text verbundener Prozesse (z. B. Herstellung der Textkohärenz) nicht möglich.

Die Überprüfung dieser Hypothesen erfolgt in Kap. 7.

6. Überlegungen zur Verbindung von C-Tests und Lesetests

Wie bereits eingangs gesagt, soll in der vorliegenden Arbeit untersucht werden, ob mit den Leistungen der getesteten Lehramtskandidaten bei der Lösung von C-Tests Aussagen zur Lesefertigkeit der Probanden verbunden sind. Obwohl in statistischen Untersuchungen im Fremdsprachenunterricht öfter hohe Korrelationen zwischen C-Test und Leseverständnistests nachgewiesen wurden (vgl. z. B. Arras et. al., 2002; Eckes & Grotjahn, 2006), ist in der Forschung weiterhin umstritten, ob diese Ergebnisse generalisierbar sind (vgl. dazu die Publikation von Grotjahn & Tönshoff (1992) und Kap. 5.1.).

Entsprechende Untersuchungen für diese Korrelationen wurden bislang nur mit Fremdsprachenlernern durchgeführt, wohingegen es in unserer Studie um erwachsene Muttersprachler mit vollentwickelten Sprachfähigkeiten geht, die differenziert werden sollen. Diese Spezifik der Adressaten hat dazu geführt, dass in den von uns entwickelten C-Tests die Wortanfangstilgung angewendet wird (vgl. Kap. 4). Durch dieses Konzept wird der Akzent der Lösungen auf die *Worterkennung* verschoben und morphologische Kenntnisse, die bei der Endtilgung mit überprüft werden, spielen bei Lösung des Tests keine Rolle mehr. Wie im vorigen Kapitel dargelegt, lassen sich für das Dekodieren und Verstehen eines Lesetextes verschiedene Prozessebenen annehmen. Deshalb sollen in der vorliegenden Studie Hypothesen in Bezug auf die Erfassung von Leseleistungen durch die C-Tests überprüft werden (vgl. dazu 6.4).

Um diese Hypothesen zu überprüfen, müssen die einzelnen Items eines C-Tests verschiedenen Prozessebenen (Wort, Satz- und Textebene) zugeordnet werden (vgl. dazu Grotjahn & Tönshoff 1992, Sigott 2004, Reitbauer 2000). Wie in den Hypothesen formuliert wurde, wird untersucht, ob zwischen den leistungsstarken und leistungsschwachen Lesern bei der Anwendung von Lesestrategien auf unterschiedlichen Ebenen Unterschiede bestehen. Es ist anzunehmen, dass einem Probanden, dem die Rekonstruktion der Lücken auf der Wortebene nicht gelingt, auch die Rekonstruktion auf der Satz- und Textebene größtenteils nicht gelingt. Löst der Proband die Lücken auf den höheren Ebenen

(Textebene), dann kann er die Lücken auf der niedrigeren Ebene (Wortebene) wahrscheinlich rekonstruieren. Das bedeutet, dass die Wortebene nicht ausreicht, um den Textsinn zu erschließen.

Es wird davon ausgegangen, dass sich Gruppen bilden lassen, die den Prozessebenen zugewiesen werden können und dass zwischen ihnen Unterschiede bestehen. Ferner wird angenommen, dass leistungsschwache Leser auf der Textebene scheitern können, da ihnen komplexe Lesestrategien fehlen. Auf der Wortebene werden die Unterschiede zwischen den Gruppen wahrscheinlich geringer ausfallen, da die Rekonstruktion auf dieser Ebene zum einen mit der Heranziehung nur des näheren Kontextes und zum anderen auch mit Vorwissen über die Grundregel bei Worttilgungen (vgl. Baur & Spettmann, 2006b) möglich ist.

Um eine Untersuchung dieser Annahmen durchzuführen, muss zunächst gezeigt werden, welche Items in einem C-Test welcher Prozessebene des Leseprozesses zugeschrieben werden können.

6.1 Können C-Tests Lesefähigkeiten messen?

Wie oben gesagt, war die Erforschung des C-Tests lange Zeit mit der Fremdsprachenforschung verbunden. Daher wurde eine Verbindung des C-Tests mit der Lesefähigkeit kritisch beurteilt. So schrieb z. B. Klein-Braley (1982), dass mithilfe der C-Tests kein Leseverständnis gemessen wird: "we could not claim that these tests can be used for measuring reading comprehension" (S. 125). Im Anschluss an die Untersuchung von Klein-Braley (1987) kamen auch Grotjahn & Tönshoff (1992) in einer eigenen Untersuchung zu dem Schluss, dass Einiges dagegen spricht, dass der C-Test Textverständnis misst. 1989 vertraten dagegen Baur & Meder (1989) für die von ihnen getestete Gruppe von *Zweitsprachenlernern* im Bereich Deutsch als Zweitsprache zum ersten Mal explizit die Meinung, dass der C-Test auch die *mit Lesen und Schreiben* verbundenen sekundären Sprachfähigkeiten testet.

Dieser Verbindung zwischen C-Test und Leseverständnis gingen Grotjahn & Tönshoff (1992) nach. Folgende Fragen standen im Zentrum ihrer Untersuchung:

Ist der C-Test ein Leseverständnistest? Und inwieweit erscheint er geeignet, das beim sinnentnehmenden Lesen erzielte Textverständnis zu messen?

Auf diese Studie wird hier näher eingegangen, da sie gegenüber vorliegenden Untersuchungen kontroverse Ergebnisse liefert. Die Autoren kamen zu dem Schluss, dass das Leseverständnis mit dem C-Test eingeschränkt gemessen werden kann. Untersucht wurde eine Gruppe von Fremdsprachenlernern in den Sprache Französisch und Italienisch. Die C-Tests wurden sowohl Anfängern als auch fortgeschrittenen Lernern zur Bearbeitung vorgelegt. Die Autoren überprüften, ob Fremdsprachenlerner nach der Rekonstruktion eines C-Test-Textes in der Lage waren, den Sinn des Textes wiederzugeben. Anschließend wurde den Studenten der unzerstörte C-Test-Text mit der Anweisung zur textnahen Übersetzung vorgelegt. In der Studie wurden miteinander verglichen: Lösungen der einzelnen C-Tests, schriftliche Wiedergabe-Protokolle und Übersetzungen der unzerstörten Texte. Die Autoren vertreten die These, dass der C-Test als integratives Verfahren zwar in gewissem Umfang *auch* Lesekompetenz messe, jedoch wegen seiner kanonischen Form mit 20 oder 25 Lücken und entsprechend kurzen Texten kein valides Instrument zur Messung des Leseverständnisses darstelle (vgl. Grotjahn & Tönshoff, 1992, S. 19). Es gebe Probanden, die bei Vorlage der unzerstörten Texte zwar ein ausreichendes Textverständnis erzielen, die den als C-Test eingerichteten Text jedoch nur sehr unbefriedigend rekonstruieren. Eine mögliche Erklärung dafür sei, dass es Lerner gäbe, die eine relativ hohe Textverstehenskompetenz erreichen, ohne jedoch z. B. über die für das Lösen von C-Tests notwendigen Morphologiekenntnisse zu verfügen. Das sei etwa dann der Fall, wenn die Fremdsprache ausschließlich in fachspezifischen Kursen mit entsprechender Ausrichtung auf die rezeptive Komponente (Verzicht auf systematische Vermittlung produktiver Grammatikkenntnisse, verbunden mit entsprechender Schreibkompetenz) gelernt worden sei (vgl. Grotjahn & Tönshoff, 1992, S. 75f.) Probanden mit einem relativ hohen C-Test-Punktwert erzielten beispielsweise ein nicht zufriedenstellendes Textverständnis. Nach Grotjahn & Tönshoff besteht eine mögliche Erklärung für diesen Sachverhalt darin, dass die entsprechenden Probanden trotz relativ hoher grammatischer Kompetenz nicht über

ausreichende Lesekompetenz verfügten (vgl. Grotjahn, 1987b, S. 240; auch Alderson & Urquhart, 1984).

Die Autoren vermuten, dass die spezifische Art der Redundanzreduzierung aufgrund des C-Test-Prinzips (hier: Wortendtilgung) möglicherweise zu einer weniger inhaltlichen, mehr formalen (syntaktisch-morphologischen) Bearbeitungsstrategie führe. Grotjahn & Tönshoff kamen zu dem Schluss, dass der C-Test in einem sehr eingeschränkten Sinne ein Lesetest zu sein scheint. Da er zusätzlich zu den für das Lesen von (fremdsprachlichen) Texte relevanten Fertigkeiten noch weitere Fertigkeiten voraussetzt, kann von einem niedrigen C-Test-Punktwert nicht notwendigerweise auf mangelnde Lesekompetenz geschlossen werden.

Beim Vorliegen eines hohen C-Test-Punktwertes kann hingegen vermutet werden, dass der Proband über eine Reihe von Fertigkeiten verfügt, die für das fremdsprachliche Lesen grundlegend sind, ohne dass er allerdings notwendigerweise eine hohe Textverstehenskompetenz besitzt (vgl. Grotjahn & Tönshoff, 1992, S. 75f.). Textverstehenskompetenz setzt vielfältige Wissensbestände voraus, da jeder Text anders ist. Sie ist ein Mittel zur Erschließung neuer Texte und schafft Zugang zu neuen Informationen, die dem Lesenden helfen, seine Lernstrategien weiterzuentwickeln. Somit stellt der C-Test nach Grotjahn & Tönshoff keine gleichwertige Alternative zu den traditionellen Leseverständnistests dar (vgl. Grotjahn & Tönshoff, 1992, S. 75).

Diese Aussage von Grotjahn & Tönshoff hat im Rahmen dieser Studie nur eine sehr eingeschränkte Gültigkeit: Zum einen haben die Autoren nur die Korrelation zu ‚klassischen‘ C-Tests mit der Wortendtilgung hergestellt. C-Tests mit Wortanfangstilgung würden mit Sicherheit zu anderen Ergebnissen führen, da hier syntaktisch-semantische Bearbeitungsstrategien aktiviert werden müssen und die Morphologie keine Rolle mehr spielt. Dieses C-Test-Format wurde von Grotjahn & Tönshoff in ihrer Untersuchung allerdings nicht erprobt, sodass die Aussagen der Autoren für uns schon daher nur eine sehr beschränkte Geltung haben (nämlich für C-Tests mit Endtilgung). Eine zweite Einschränkung bezieht sich darauf, dass die Überprüfung bei Fremdsprachenlernern vorgenommen wurde. Für Zweitsprachenlerner und Muttersprachenlerner können völlig andere

sprachliche Voraussetzungen und damit verbundene Interkorrelationen zwischen den sprachlichen Fertigkeiten angenommen werden.

Diese Meinung wird in neuerer Zeit von Baur, Grotjahn & Spettmann (2006) vertreten. Sie behaupten, dass C-Tests als Instrument zur Überprüfung der mit Lesen und Schreiben verbundenen Fertigkeiten einsetzbar sind und diese (auch) das Textverständnis messen (S. 7f.). Im Bereich DaZ sehen Baur & Spettmann (2010) den C-Test

„...in erster Linie als Instrument an, mit dem Lesekompetenz und Textverständnis überprüft werden können. In diesem Zusammenhang wird der C-Test als Test- und Förderinstrument für das Lesen angeboten.“ (Baur & Spettmann, 2010, S. 95)

Einen Schwerpunkt legen die Autoren auf den Übergang von der Primarstufe in die Sekundarstufe und auf den Beginn der Sekundarstufe. Sie betonen, dass es auf dieser Stufe besonders wichtig sei, zu erkennen, ob Lesekompetenz ausreichend ausgebildet worden ist oder Fördermaßnahmen notwendig sind (vgl. dazu auch Baur & Spettmann, 2007, und Baur & Spettmann, 2009).

Weitere Einsätze von C-Tests als Instrument zur Überprüfung der Lesefähigkeit bestätigten das Interesse von Sprachwissenschaftlern und Testentwicklern an dem Testformat als Erweiterung zu klassischen Leseverständnistests. Zur Ermittlung der allgemein- und fachsprachlichen Lesekompetenz im Berufskolleg setzten Baur, Mashkovskaya & Spettmann (2010) adaptierte, adressatenspezifische C-Tests ein. Im Rahmen dieser Untersuchung wurde überprüft und bestätigt, dass Berufsschüler mit einer gering ausgebildeten allgemeinen Lesekompetenz auch schwächere Ergebnisse im fachspezifischen Leseverständnis erzielen und umgekehrt. Diese Pilotstudie zeigte, dass durch die Ermittlung der Worterkennungs-Werte (WE-Werte) aussagekräftige Daten bezüglich des allgemeinen und fachsprachlichen Textverständnisses und der Lesekompetenz der Schüler geliefert werden. Als sinnvoll hat sich erwiesen, das Verständnis von Fachtexten nicht abzukoppeln, sondern in Kombination mit dem allgemeinsprachlichen Textverständnis (und somit auch das Verhältnis zwischen beiden Bereichen) zu überprüfen. Die Erhebung der Differenzwerte¹⁰⁰ (das

¹⁰⁰ Zur Ermittlung, Berechnung und Interpretation des Differenzwertes vgl. Baur & Spettmann 2010.

Verhältnis zwischen WE- und R/F-Werten) liefert wichtige Informationen für die sprachliche Förderung der Berufsschüler.

Auf die Unterschiede zwischen der getesteten Gruppe der Berufsschüler und der vorliegenden Studie wird hier hingewiesen. Die Untersuchungen für die C-Tests wurden von Baur und Spettmann nicht mit erwachsenen Sprechern durchgeführt, sodass es sich nicht um dieselbe Adressatengruppe handelt wie in der vorliegenden Studie. Des Weiteren ermitteln Baur & Spettman durch die Berechnung der WE-Werte und des Differenzwertes (DF-Wertes) Daten, die nach ihrer Argumentation Informationen über das Textverständnis der Probanden liefern; doch sie haben ihre Ergebnisse für die Tests nicht mit anderen Lesetests abgeglichen.

Im Rahmen der vorliegenden Studie wird nun erstmals der Versuch unternommen, C-Tests und Tests zur Überprüfung der Lesekompetenz systematisch miteinander zu vergleichen. Angenommen wird, dass für die erfolgreiche Rekonstruktion von C-Tests mit Wortanfangstilgung Lesestrategien notwendig sind, die für die Lösung anderer Lesetests ebenfalls eine Grundlage bilden. Diese Lesestrategien können verschiedenen Leseprozessebenen zugeordnet werden (vgl. Kap. 6.2), sodass ein Einblick in die Rekonstruktion/Bearbeitung von Lücken geschaffen werden kann. Wird der C-Test erfolgreich rekonstruiert, wird davon ausgegangen, dass der Lesende auch den Textinhalt verstanden hat und über das notwendige Text- bzw. Leseverständnis verfügt. Kann der Lesende nur sehr wenige Lücken ausfüllen und lässt er die anderen unbearbeitet, scheitert er möglicherweise, weil er eine der notwendigen Lesestrategien nicht anwenden kann. Auf welches Niveau der Lesekompetenz durch solche Beobachtungen geschlossen werden kann, soll zu einem späteren Zeitpunkt betrachtet werden.

Damit soll im Rahmen der vorliegenden Studie untersucht werden, inwieweit C-Tests Lesefähigkeiten der Studierenden abbilden können. Dafür werden zum Vergleich mit den C-Tests verschiedene Testformate zur Überprüfung der Lesefähigkeiten eingesetzt (vgl. Kap. 5, Kap. 7). Durch korrelative Berechnungen

soll im Folgenden kontrolliert werden, inwieweit die verwendeten Lesetests mit den C-Tests ähnliche Lesekompetenzkomponenten messen. Bestehen mittlere oder hohe Korrelationen, wird angenommen, dass beide Testverfahren Teilaspekte der Lesekompetenz auf ähnliche Weise erfassen. Besteht kein Zusammenhang zwischen den jeweiligen Testinstrumenten, dann messen die verwendeten C-Tests und der respektive Lesetest entweder a) unterschiedliche Konstrukte oder Teilaspekte der Lesekompetenz oder b) die beiden Testformate messen die Lesekompetenz auf unterschiedliche Weise.

6.2 Prozessebenen bei der Bearbeitung von C-Tests

Bereits in den achtziger und neunziger Jahren wurde versucht, die C-Test-Lücken zu kategorisieren, um Beziehungen zwischen einzelnen Lücken zu analysieren (vgl. Grotjahn, 1984; Sigott 1998; Stemmer, 1991). Die Ergebnisse neuerer Untersuchungen (vgl. Reitbauer, 2000; Harsch & Hartig, 2010) bezeugen, dass der C-Test nicht nur, wie vorher angenommen, in einem Mikrokontext, sondern auch in einem Makrokontext messen kann. Beim Mikrokontext handelt es sich um hierarchieniedrige Prozesse oder Lösungsstrategien, die verwendet werden, wenn zum Beispiel nur auf Wort- oder Satzebene gearbeitet wird (vgl. Prozessmodelle von van Dijk & Kintsch, 1985 und Christmann & Richter 2002 in Kap. 2.2.3). Auf der Wort- und Satzebene werden in der Regel syntaktisch-morphologische Fertigkeiten aktiviert. Beim Makrokontext dagegen geht es um hierarchiehöhere Prozesse und Strategien, die satzübergreifend angewendet werden und die bei der Bearbeitung Wissen über Textstruktur, Inhalt und Weltwissen aktivieren.

Die Kategorisierung der Lücken im C-Test ist nicht immer eindeutig, da Lücken beispielsweise verschiedenen Kategorien gleichzeitig zugeordnet werden können. Reitbauer (2000), die in ihrer Studie C-Tests im Englischen untersuchte, die auf Grundlage fachspezifischer Abstracts aus dem Bereich Psychologie entwickelt wurden, beschreibt, wie die Kategorienzuordnung einzelner Lücken stattfand:

„Sobald es plausibel schien anzunehmen, dass die Lösung des Items ohne die Miteinbeziehung eines über die Satzgrenze hinausgehenden Kontextes erfolgen kann, wurde davon ausgegangen, dass dieses Item Fertigkeiten auf den niedrigen Verarbeitungsebenen misst.“ (Reitbauer, 2000, S. 227)

Der höheren Verarbeitungsebene wurden solche Items zugeordnet, von denen angenommen wurde, dass der Proband zu ihrer Lösung mit großer Wahrscheinlichkeit den Makrokontext einbeziehen musste. Der Begriff Makrokontext wird in dieser Untersuchung über die Satzgrenze definiert. Sobald die Lösung eines Items erfordert, über die Satzgrenze hinauszublicken und die richtige Interpretation des vorangegangenen oder nachfolgenden Satzes voraussetzt, gilt es als Item, das Fähigkeiten auf höheren Verarbeitungsebenen misst (vgl. Reitbauer, 2000, S. 220).

In welcher Reihenfolge diese Prozesse sich vollziehen, ist umstritten (vgl. Kap. 2.2). Die meisten Linguisten gehen von einem Modell aus, bei dem aufsteigend zwischen grapho-phonischer, lexikalisch-formaler, syntaktischer und semantischer Ebene unterschieden wird. Da der Leseprozess die Aktivierung verschiedener Prozesse erfordert (sowohl hierarchieniedriger als auch hierarchiehöherer) werden von den Lesern beide Verarbeitungsstrategien verwendet.

Eine Lückenkategorisierung für die englischen C-Tests wurde im Rahmen der DESI-Studie auch von Harsch & Hartig (vgl. Harsch & Schröder, 2008, S. 149ff.) vorgenommen. In der DESI-Studie wurde der C-Test auf zwei Ebenen charakterisiert: zum einen auf der Textebene, zum anderen auf der Itemebene (Lückenebene).

Der Itemfokus wurde auf den ganzen Text bezogen in drei Stufen erfasst:

1. Fokus auf Basisstrukturen und Grundwortschatz; wenige Lücken, die auf textgrammatische Phänomene fokussieren, wobei es sich um einfache Strukturen handelt; weitgehende Vermeidung von Lücken, die zur Schließung ein komplexes Zusammenspiel linguistischen Wissens erfordern.
2. Fokus auf Strukturen und Wortschatz mittleren Schwierigkeitsgrades; einige Lücken, die auf textgrammatische Phänomene auch über Basisstrukturen hinaus fokussieren; einige Lücken, die auf das komplexe Zusammenspiel des linguistischen Wissens fokussieren.

3. Ausgeglichener Anteil der Lücken mit Fokus auf komplexe Strukturen und Spezialwortschatz; Fokus auf komplexe textgrammatische Phänomene; Fokus auf komplexes Zusammenspiel des linguistischen Wissens.

Bei Muttersprachlern ist eine solche Zuordnung aus zwei Gründen schwierig: Zum einen ist die Kategorisierung bezüglich des Wortschatzes problematisch, weil nicht klar ist, was als Maß für die Bestimmung der Schwierigkeitsgrade dienen soll. Im Fremdsprachenbereich orientiert die Analyse sich häufig an den GER-Niveau-Stufen, die auch die Wortschatzbreite und das Themenspektrum beschreiben.

Bei schriftlichen Texten, die Studierende selbst produzieren, käme eine solche Wortschatzanalyse in Frage, falls man die Qualität der Texte mitbewerten würde, da nicht jeder Text, der einen breiten Wortschatz zeigt, bereits als qualitativ gut eingeschätzt werden kann. Für die Verwendung der Kategorisierung beim C-Test auf der Wortschatzebene, wo die Kontextgrenzen in der Regel sehr eng sind, eignet sich eine derartige Analyse nicht. Zum anderen verrät die Breite des Wortschatzes nichts darüber, ob der Textkontext im Rahmen eines C-Tests verstanden wurde: Es kann durchaus vorkommen, dass die Lücke durch ein selten gebrauchtes Wort vervollständigt wird, das dem Satz- oder Gesamtkontext jedoch widerspricht, was verdeutlicht, dass der Proband zwar belesen ist, den Inhalt jedoch stellenweise offenbar nicht begreift.

An dieser Stelle wird entschieden, die Lücken anders zu kategorisieren, nämlich in Bezug auf die Leseleistung und die damit verbundenen Prozesse, die sich durch eine erfolgreiche bzw. weniger erfolgreiche Rekonstruktion beobachten lassen.

6.3 Kategorisierung der Lücken

In Anlehnung an die in Kap. 4 beschriebenen Leseprozessmodelle und unter besonderer Berücksichtigung des Lesemodells von Christmann & Richter (2002) werden für meine Studie folgende Kategorien formuliert, denen einzelne Lücken im C-Test zugeordnet werden. Die Aufteilung erfolgt in drei große Bereiche: Wort-, Satz- und Textebene. Um die Zuordnung zu der jeweiligen Ebene zu

kennzeichnen, werden Abkürzungen verwendet. Die Kategorisierung der Lücken erfolgt nach folgendem Schema:

1. Wortebene (W-Ebene)

- Die Bearbeitung der Lücke auf der Wortebene verlangt nur die Miteinbeziehung des *Kontextes*¹⁰¹ *im engeren Sinne*, d. h. für die Rekonstruktion der Lücke sind nahe liegende Wörter von Bedeutung. In der Regel sind es ca. 1-2 Wörter vor und nach der Lücke. Folgende Lücken sind ohne Einbeziehung der syntaktischen Struktur des Satzes zu rekonstruieren, z.B. freue mich, des Gesprächs, zur Not, in dieser Gegend.

2. Satzebene (S-Ebene)

- Die Bearbeitung der Lücke auf dieser Ebene verlangt die Verwendung von Kontext im weiteren Sinne, d. h. für die Rekonstruktion der Lücke sind Wörter notwendig, die im Rahmen des Satzes vorkommen. Als erweiterter Kontext wird hier der komplette Satz definiert. Auf dieser Ebene spielen syntaktische Analysen eine große Rolle.

- Die Bearbeitung der Lücke auf dieser Ebene verlangt die Verwendung von komplexeren Strategien, zum Beispiel, wenn zusammengehörende Teile rekonstruiert werden müssen und diese im Satz nicht zusammen stehen, sondern durch mehrere Wörter (ab)getrennt sind, z.B.: *Offensichtlich treffen Eltern nach der Grundschulzeit ihre Schulwahl für ihre Kinder vor allem in Abhängigkeit vom eigenen Schulabschluss.*

- Die Bearbeitung der Lücke auf dieser Ebene verlangt die Verwendung von logischen Bezügen zu anderen Sätzen, etwa, wenn *deiktisch-anaphorische* Verknüpfungen (*er, der Professor; sie, eine sehr junge Mutter* etc.) mithilfe des vorangegangenen oder des nachfolgenden Satzes rekonstruiert werden müssen.

¹⁰¹ Im Folgenden wird zwischen Ko- und Kontext unterschieden, da dies für die Kategorisierung und Analyse der Lücken wichtig ist.

Dieser Kategorie sind auch Lücken zuzuordnen, die in längeren, verschachtelten Sätzen vorkommen, z.B.: *Lisa ist unsere neue Mitarbeiterin, die seit einem Monat bei uns in der Firma tätig ist. Geboren in Südtirol, in einem kleinen Dorf namens Ölde, wuchsen ihre zwei kleinen Brüder auf, denen sie eine ältere Schwester war.*

3. Textebene (T-Ebene)

- Die Bearbeitung der Lücken auf der Textebene verlangt die Verwendung satzübergreifender Strategien – zum Beispiel, wenn ein Bezug zum Textkontext und über den Textkontext hinaus hergestellt werden soll. Für die Rekonstruktion dieser Lücken ist das Erkennen von Beziehungen zwischen den Sätzen notwendig. Manche C-Test-Items bilden ‚Lücken‘ im Verstehensprozess, was zu Missverständnissen und Sinnverschiebungen führen kann. Hier wird die Herstellung von Kohärenz sowohl auf Satz- als auch auf Textebene vorausgesetzt. An dieser Stelle kann sowohl der Bezug auf die Überschrift des Textes als auch auf das Thema allgemein nützlich sein. Vorwissen und Hintergrundwissen spielen dabei eine wichtige Rolle; die Lücken können nicht mehr durch semantisch-syntaktische Hinweise im Text rekonstruiert werden. Beispiel: Die Studenten verwenden viel Mühe und Zeit für die Bewältigung komplexer Aufgaben. Hier sind die Probanden gezwungen, höhere, komplexere Verarbeitungsstrategien anzuwenden. Auf dieser Ebene werden textgestützte Hypothesen überprüft, um zu bestimmen, ob der Text durch die Vervollständigung der Lücken kohärent bleibt (globale Kohärenzbildung, Verwendung des Situationsmodells, Vervollständigung von Propositionen).

6.4 Kategorisierung der Lücken an einem C-Test-Beispiel

Im Folgenden wird anhand eines C-Test-Beispiels (Teilttest 1 aus der Testbatterie)¹⁰² die Kategorisierung der Lücken vorgenommen, um die Zuordnung zu verschiedenen Ebenen zu veranschaulichen. Für die Analyse wurden die Daten der Gesamtstichprobe (N = 887) verwendet. Die

¹⁰² Komplette C-Testbatterie befindet sich im Anhang A, Teil 2.1.

Gesamtgruppe der Studierenden wurde in zwei Untergruppen aufgeteilt, die Gruppe der leistungsstarken und leistungsschwachen Leser. Die Aufteilung erfolgte in Bezug auf den Referenznormwert, der für die Lehramtsstudierende für die vorliegenden C-Tests festgelegt wurde (vgl. dazu Kap. 4.9, Kap. 7.2). Lag die Leistung eines Probanden unter dem Referenznormwert, wurde er der Gruppe der leistungsschwachen Leser zugeordnet. Lag sie über dem Referenznorm, erfolgte eine Zuordnung zur Gruppe der leistungsstarken Leser. Nach der Aufteilung der Gruppe in Bezug zum Referenznormwert besteht die Gruppe der leistungsstarken Leser aus 592 (66,7 %) Probanden und die der Leistungsschwachen aus 276 (33,3 %) Probanden.

Einzelne Pfeile in den folgenden Abbildungen sollen die Beziehungsrichtung (z. B. vorwärts oder rückwärts) und Beziehungsebene (Wort-, Satz oder Textebene) illustrieren.

Abb. 7: Kategorisierung der Lücken auf der Wortebene

T1: Auf dem Weg zum zweigliedrigen Schulwesen

Seit Kurzem wird diskutiert, ob das jetzige dreigliedrige Schulwesen durch ein zweigliedriges System ersetzt werden sollte. Dafür (L1) gibt es (L2) folgenden Grund: Immer mehr Eltern wünschen (L5) sich bessere Abschlüsse für ihre Kinder. Diesen Wunsch assoziieren sie bundesweit mit (L11) der Wahl weiterführender Schulen.

(L13) Der Wandel ist die logische Konsequenz (L16) des Zusammenwirkens mehrerer Faktoren. Jenseits offizieller Verlautbarungen, wonach Schulentscheidungen allein nach Maßgabe von Leistung, Eignung und Fähigkeit erfolgen, treffen Eltern nach der Grundschulzeit ihre Schulwahl vor allem in Abhängigkeit vom eigenen Schulabschluss. Offensichtlich gilt, dass Kinder in einen Bildungsgang einmünden sollen, der einen besseren Schulabschluss verheißt, als ihn die Eltern erworben haben.

Im Folgenden werden exemplarisch einige Lücken in Bezug auf ihre Rekonstruktion hin beschrieben. Wie bereits bei der Kategorisierung der Lücken erwähnt (Kap. 6.3), verlangt die Bearbeitung auf der Wortebene nur die Miteinbeziehung des Kontextes im engeren Sinne (z. B. nahliegender Wörter). Das verdeutlichen farbliche Markierungen und Pfeile (Abb. 8), welche die Richtung der Wortbeziehungen zeigen.

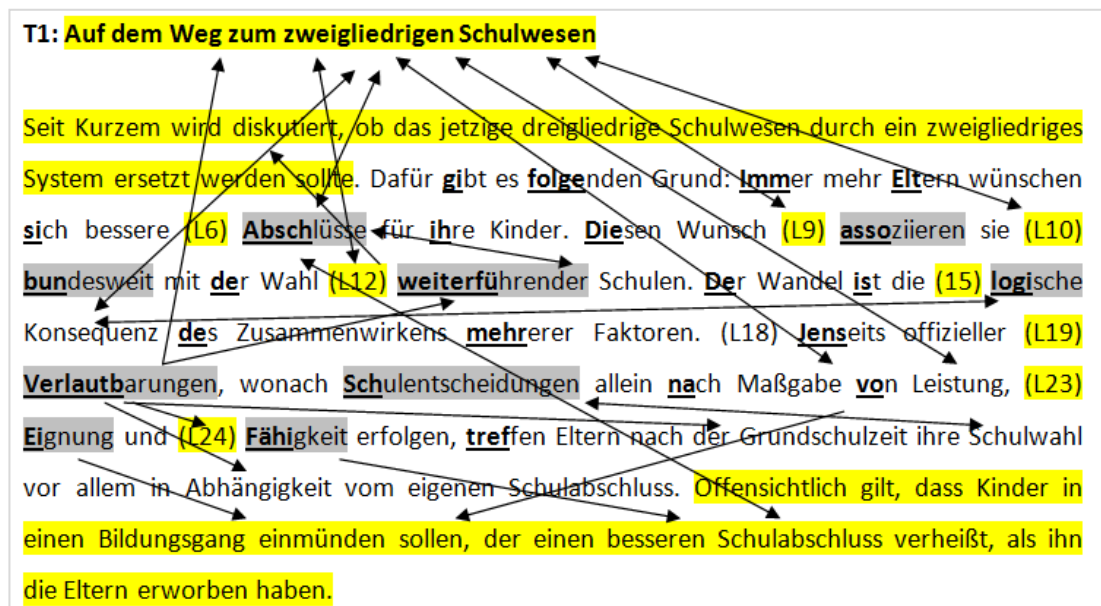
Abb. 8: Kategorisierung der Lücken auf der Satzebene**T1: Auf dem Weg zum zweigliedrigen Schulwesen**

Seit Kurzem wird diskutiert, ob das jetzige dreigliedrige Schulwesen durch ein zweigliedriges System ersetzt werden sollte. Dafür gibt es folgenden Grund: (L3) Immer mehr (L4) Eltern wünschen sich bessere Abschlüsse für (L7) ihre Kinder. (L8) Diesen Wunsch assoziieren sie bundesweit mit der Wahl weiterführender Schulen. Der Wandel (L14) ist die logische Konsequenz des Zusammenwirkens (L17) mehrerer Faktoren. Jenseits offizieller Verlautbarungen, wonach Schulentscheidungen allein (L20) nach Maßgabe (L21) von Leistung, Eignung und Fähigkeit erfolgen, (L25) treffen Eltern nach der Grundschulzeit ihre Schulwahl vor allem in Abhängigkeit vom eigenen Schulabschluss. Offensichtlich gilt, dass Kinder in einen Bildungsgang einmünden sollen, der einen besseren Schulabschluss verheißt, als ihn die Eltern erworben haben.

Bei der Rekonstruktion der Lücken auf der Satzebene reicht die Wortebene nicht mehr aus; hier ist die Einbeziehung des ganzen Satzes notwendig und wichtig. Liest der Proband den Satz z. B. bei der Lücke L14 (vgl. Abb. 8) bis zum Ende, kann er aufgrund seiner Kenntnisse über den syntaktischen Satzbau sehen, dass an der Stelle, wo die Lücke eingebaut ist, ein Verb in der dritten Person Singular eingesetzt werden muss. Die Lücken L20, L21 und L25 lassen sich auch innerhalb des Satzes rekonstruieren. Die Präpositionen *nach* und *von* gehören zum Verb *erfolgen* und zum Substantiv *Maßgabe*. Um die Lücken füllen zu können, muss syntaktisches Wissen über die Struktur dieses komplexen Satzes mit einem eingeschobenen Nebensatz aktiviert werden. Gelingt dies den Probanden, können sie die Lücken mit entsprechenden Präpositionen füllen.

Die Lücke L25 mit dem Verb *treffen* lässt sich rekonstruieren, wenn man im Nebensatz den zweiten Teil dieses Ausdrucks findet: *ihre Schulauswahl treffen*. Da erfahrungsgemäß nicht alle Probanden den letzten Satz bis zum Ende lesen (weil dieser tilgungsfrei im Text steht), kann es dazu kommen, dass die Lücke falsch ausgefüllt wird, z. B. wird statt *treffen* das Verb *helfen* eingesetzt.

Abb. 9: Kategorisierung der Lücken auf der Textebene



Auf der Textebene geht es um hierarchiehöhere Prozesse und Strategien, die satzübergreifend angewendet werden und die Wissen über die Textstruktur, den Inhalt und das Weltwissen aktivieren. Die Lücken können nicht mehr durch semantisch-syntaktische Hinweise im Text rekonstruiert werden. Die Probanden sind gezwungen, komplexere Verarbeitungsstrategien anzuwenden. Die im Text gezeichneten Pfeile (Abb. 9) zeigen verschiedene, über die Satzgrenze ausgehende Bezüge zwischen den Lücken und den Textteilen, die für die Lösung der Lücken wichtige Informationen enthalten.

Die dargestellten Abbildungen zeigen, dass die Kategorisierung der Lücken wichtige Ansatzpunkte bietet, die über Richtig/Falsch-Auswertung der Lücken hinausgehen und die es ermöglichen, allgemeine Annahmen in Bezug auf die Bearbeitung auf verschiedenen Prozessebenen zu stellen und zu überprüfen.

Durch die Kategorisierung der Lücken soll u. a. auch untersucht werden, ob und auf welchen Prozessebenen Unterschiede zwischen den leistungsstarken und leistungsschwachen Lesern bestehen.

In Anlehnung an das Lesemodell von Richter & Christmann (2002) und die Untersuchungen von Reitbauer (2000) über fachspezifische C-Tests im Englischen wird angenommen, dass zwischen den oben genannten Gruppen Unterschiede bestehen. Ein größerer Abstand würde vermutlich auf der Satz- und/oder

Textebene bestehen, da sie komplexere Lese- und Rekonstruktionsstrategien erfordern. Wahrscheinlich werden leistungsschwache Leser auf diesen Ebenen scheitern, da ihnen komplexe Lesestrategien fehlen oder nicht automatisiert zum Einsatz kommen. Auf der Wort- und Satzebene werden die Unterschiede zwischen den Gruppen voraussichtlich geringer ausfallen, da die Rekonstruktion auf dieser Ebene zum einen mit der Heranziehung nur des näheren Kontextes und zum anderen mit Vorwissen über die Grundregel bei Worttilgungen möglich ist (vgl. Baur, Grotjahn & Spettmann, 2006, S. 12).

6.5 Hypothesen zur Differenzierung zwischen den Probanden auf den Prozessebenen

Im Folgenden werden drei Hypothesen in Bezug auf die Kategorisierung der Lücken und der Prozessebenen untersucht.

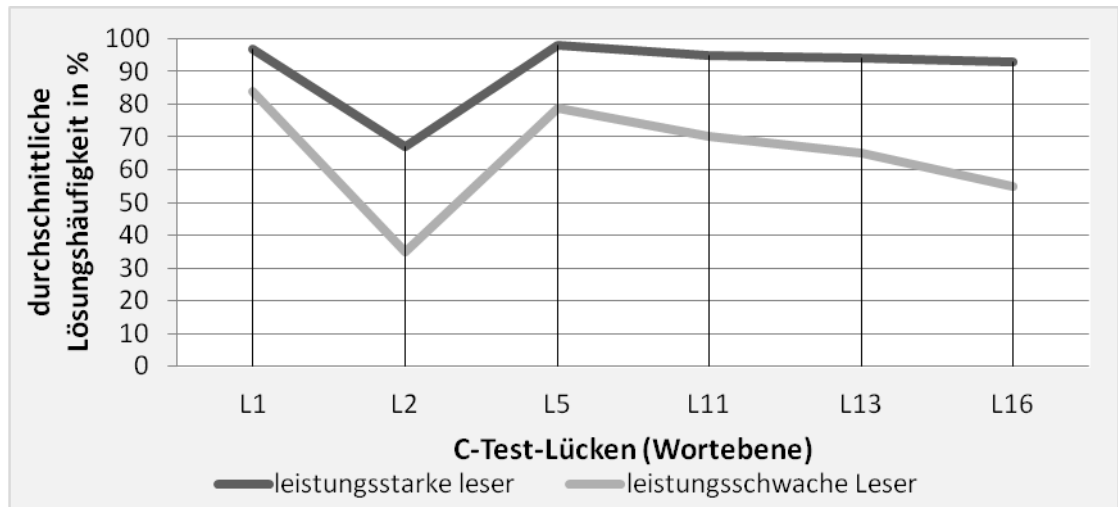
Hypothese 1: *Studierende, die gute C-Test-Ergebnisse erzielen, lösen Items, die auf einer höheren Ebene des Leseprozesses (der Textebene) liegen.*

Hypothese 2: *Studierende, die weniger gute C-Test-Ergebnisse erzielen, lösen weniger Items, die auf einer höheren Ebene (der Textebene) liegen.*

Hypothese 3: *Die Lösung der Items auf einer höheren Ebenen (der Textebene) impliziert die Lösung der Items auf einer niedrigeren Ebene.*

Analysiert man die Ergebnisse der Studierenden auf der Wortebene in Bezug auf die oben formulierten Hypothesen, müsste dies bedeuten, dass zwischen den leistungsstarken und leistungsschwachen Studierenden geringe Unterschiede bestehen, da ein leistungsschwächerer Leser in der Lage sein sollte, Strategien auf Wortebene anzuwenden. Hier operiert er mit dem näheren Kontext.

Abb. 10: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der Wortebene

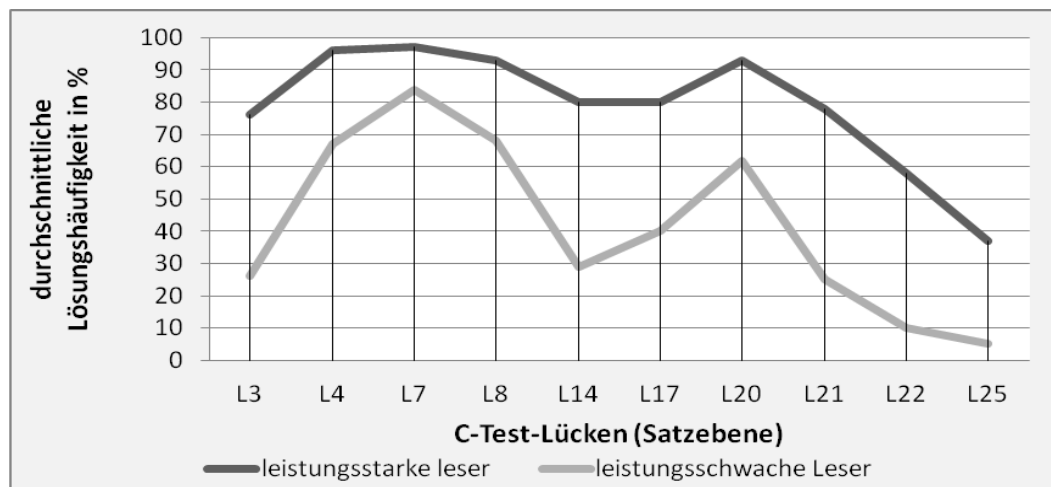


Wie Abb. 10 verdeutlicht, bestehen zwischen leistungsstarken und leistungsschwachen Lesern geringe Differenzen. Die Abstände in der erreichten durchschnittlichen Leistung variieren zwischen 13 bis 38 Punktwerten. Der größte Abstand wird bei der Lücke L16 erreicht. In der Lücke 16 sollten Studierende den Artikel *des* (Zusammenwirkens) rekonstruieren. Dass es manchen Studierenden nicht gelingt, diese Lücke zu vervollständigen, kann so interpretiert werden, dass der Genitiv als Kasus im Deutschen heutzutage weniger verwendet wird als der Dativ. Die Leistungsunterschiede zwischen der Gruppe der leistungsstarken und leistungsschwachen Probanden auf Wortebene sind signifikant (vgl. Daten im Anhang A, Teil. 3.1).

Die Hypothese Nr. 1 wurde bestätigt.

Nach den oben formulierten Hypothesen müssten auf der Satzebene größere Leistungsdifferenzen zwischen leistungsstarken und leistungsschwachen Probanden bestehen, da auf der Satzebene die Aktivierung anderer Strategien notwendig ist. Syntaktische Zusammenhänge können hier eine größere Rolle spielen als auf der Wortebene. Kann der Proband auf der Wortebene Lücken ohne Einbeziehung des gesamten Satzes rekonstruieren, so gelingt es ihm auf der Satzebene nicht mehr (vgl. Daten im Anhang A, Teil 3.1).

Abb. 11: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der Satzebene

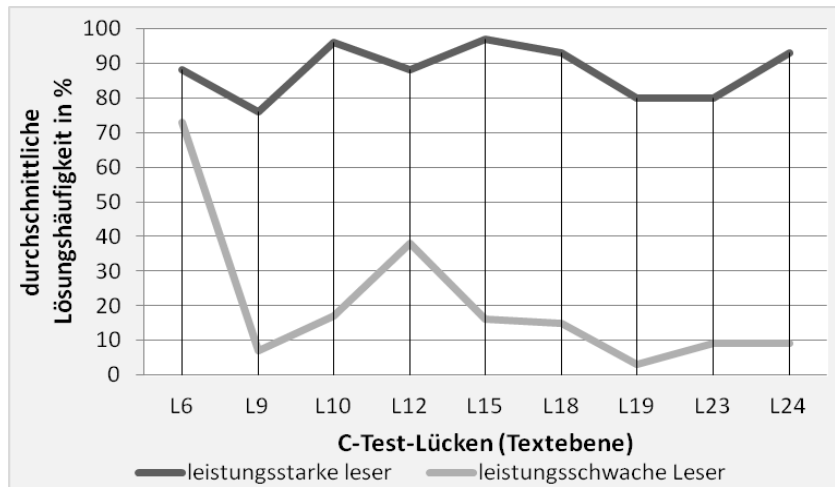


Wie die Abb. 11 zeigt, bestehen die Differenzen zwischen den beiden Gruppen auch auf der Satzebene; bei manchen C-Test-Lücken sind die Unterschiede stärker zu erkennen. Die Leistungsdifferenzen vergrößern sich in einzelnen Fällen (vgl. L3, L14, L17, L21, L22) und variieren zwischen 15 (vgl. L7) und 60 Punkten (vgl. L14). Die Leistungsunterschiede zwischen der Gruppe der leistungsstarken und leistungsschwachen Probanden auf der Satzebene sind signifikant (vgl. Daten im Anhang A, Teil. 3.1).

Die Hypothese Nr. 2 wurde bestätigt.

Betrachtet man die durchschnittliche Lösungshäufigkeit in beiden Gruppen auf der Textebene (s. Tabelle 12), so sind diese Abstände noch größer, was darauf schließen lässt, dass es den leistungsschwächeren Probanden hier an bestimmten Lesestrategien fehlt, was teilweise auch zu Problemen beim Textverständnis führen kann. Diese Differenzen lassen sich mit dem C-Test identifizieren (vgl. Berechnungen im Anhang A, Teil 3.1).

Abb. 12: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der Textebene



Die Grafik zur Rekonstruktion der Lücken auf verschiedenen Ebenen zeigt folgende Tendenz: Bei der Lösung aller Lücken sind die Ergebnisse der stärkeren Gruppe besser. Die Abstände vergrößern sich in dem Maße, wie die verschiedenen angenommenen Verarbeitungsebenen der Lücken analysiert werden. Die Lücken, deren Bearbeitung komplexere Strategieranwendungen verlangt, werden von leistungsschwachen Probanden im Durchschnitt (wie erwartet) seltener gelöst als von leistungsstarken Probanden. Auf der Textebene sind die Differenzen so groß, dass sie fast 80 Punkte ausmachen. Die hierarchiehöhere Prozessebene setzt eine intensivere Auseinandersetzung mit dem Text voraus und zeichnet lesestarke Probanden aus. Erreicht der Proband eine niedrige Punktzahl, so sind es in der Regel die Lücken, die hierarchieniedrigen Verarbeitungsstrategien zugeordnet werden (Wort- und Satzebene). Kann der Proband hierarchiehöhere Verarbeitungsstrategien (Textebene) verwenden, dann kann er, wie es sich herausgestellt hat, die hierarchieniedrigen Strategien automatisch anwenden. Die Leistungsunterschiede zwischen der Gruppe der leistungsstarken und leistungsschwachen Probanden auf der Textebene sind signifikant (vgl. Berechnungen im Anhang A, Teil 3.1).

Die Hypothese Nr. 3 wurde bestätigt.

Operiert man bei der Zuordnung der Lücken verschiedener Kategorien nicht mit drei Prozessebenen, sondern nur mit zwei, nämlich einer hierarchieniedrigen und einer hierarchiehöheren, wie es Richter & Christmann (2002) in ihrem

6. Verbindung von C-Tests und Lesetests

Lesemodell vorschlagen, so lassen sich die C-Test-Lücken auf der Wort- und Satzebene den hierarchieniedrigeren Kategorien oder Prozessebenen und C-Test-Lücken auf der Textebene den hierarchiehöheren Prozessebenen zuordnen. Die hierarchieniedrige Prozessebene stellt, wie oben Abb. 13a verdeutlicht, im Durchschnitt keine größere Schwierigkeit für schwache Leser dar. Sie verwenden Strategien zur Rekonstruktion der Lücken auf der Wortebene häufiger, da diese Ebene eine einfachere Basis für das Verstehen der Texte bildet als Strategien auf den höheren Ebenen (Satz- und Textebene).

Wenn man die Wort- und Satzebenen nach Richter & Christmann (2002) als untere Ebene zusammenfasst und mit der Textebene vergleicht, ergibt sich folgendes Bild:

Abb. 13a: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der hierarchieniedrigen Prozessebene

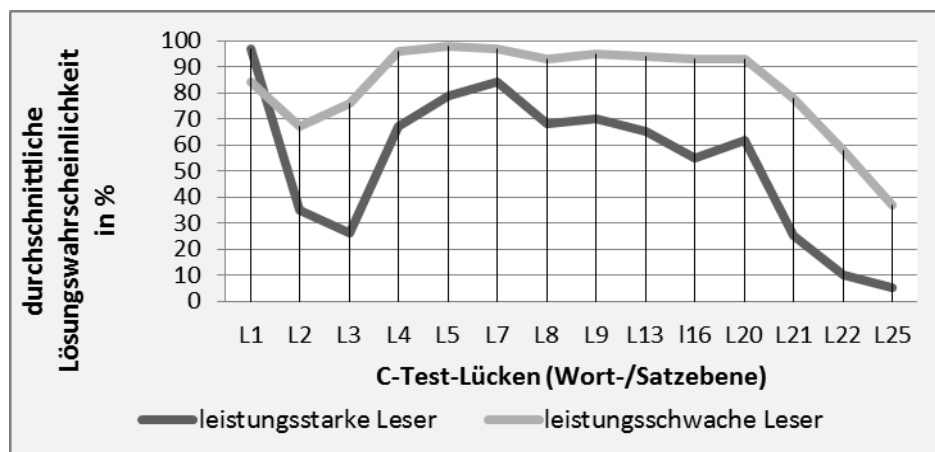
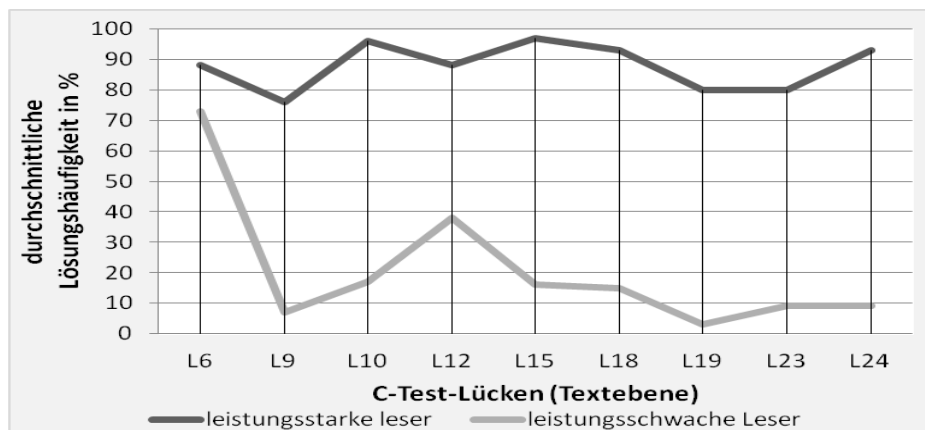


Abb. 13b: Durchschnittliche Leistung der leistungsstarken und -schwachen Leser auf der höheren Prozessebene



Vergleicht man beide Tabellen (13a und 13b) miteinander, so werden die Leistungsunterschiede zwischen den Gruppen deutlich. Betrachtet man die Lücken genauer, könnte man vermuten, dass die Lücke 6 von den leistungsschwachen Probanden *zufällig* gelöst wurde oder dadurch, dass der davorstehende Satz tilgungsfrei zur Verfügung steht und die *Lücke 6* sich ziemlich am Anfang im C-Test befindet, sodass die Rekonstruktion einfacher ist als bei anderen Lücken, die der Textebene zugeordnet sind.

Hierarchiehöhere Prozessebenen setzen eine intensivere Auseinandersetzung mit dem Text voraus und zeichnen lestarke Probanden aus (vgl. Abb. 13a und 13b). Erreicht der Proband eine niedrige Punktzahl in einem C-Test, sind häufiger Lücken rekonstruiert, die den hierarchieniedrigen Verarbeitungsstrategien zugeordnet werden (Wort- und Satzebene). Sind die C-Test-Lücken gelöst, für deren Bearbeitung hierarchiehöhere Verarbeitungsstrategien auf der Textebene verwendet werden müssen, dann sind die Lücken auf der Wort- und Satzebene in der Regel auch rekonstruiert.

Zusammenfassend lässt sich festhalten, dass

1. sich durch den Einsatz adressatenspezifischer C-Tests verschiedene Prozessebenen beobachten lassen, die beim Lesen eine wichtige Rolle spielen,
2. durch den Einsatz adressatenspezifischer C-Tests zwischen leistungsstarken und leistungsschwachen Lesern auf verschiedenen Ebenen differenziert werden kann,

3. sich durch den Einsatz adressatenspezifischer C-Tests auf den hierarchieniedrigeren Ebenen (Wort- und Satzebene) zwischen den Gruppen Leistungsunterschieden beobachten lassen,
4. sich durch den Einsatz adressatenspezifischer C-Tests auf der hierarchiehöheren Ebene (Textebene) zwischen den Gruppen sehr große Leistungsunterschiede beobachten lassen,
5. sich durch den Einsatz adressatenspezifischer C-Tests beobachten lässt, dass zwischen den Leistungen der leistungsstarken und leistungsschwachen Gruppe auf allen Ebenen signifikante Unterschiede bestehen (vgl. dazu Berechnungen in Anhang A, Teil 3, Tabellen 1 bis 12).

Die oben formulierten Hypothesen in Bezug auf verschiedene Prozessebenen und Leistungen der Probanden können danach als bestätigt gelten.

Durch die Kategorisierung der Lücken konnte gezeigt werden, auf welchen Prozessebenen der C-Test bei Muttersprachlern misst. Da der C-Test nicht alle Determinanten des Lesemodells von Artelt et al. (2007) abdecken kann (vgl. Kap. 2.2.1, Kap. 2.3), werden im Rahmen dieser Studie auch Korrelationen mit weiteren Lesetests (vgl. Kap. 6) hergestellt, und durch entsprechende Vergleiche (vgl. Kap. 7.3) werden Überschneidungspunkte zwischen den Lesetests und den C-Tests untersucht. Bestehen mittlere oder hohe Korrelationen, wird angenommen, dass beide Testverfahren ähnliche Teilaspekte der Lesekompetenz erfassen. Besteht kein Zusammenhang zwischen beiden Testinstrumenten, muss angenommen werden, dass der C-Test und der damit verglichene Lesetest unterschiedliche Konstrukte messen.

In der aktuellen Leseforschung hat sich weitgehend die Auffassung durchgesetzt, dass der Leseprozess ein Wechselspiel zwischen hierarchieniedrigen und hierarchiehöhen Prozessen ist. Allerdings gelingt die Anwendung beider Prozesse nicht allen Lesenden, wie sich in unserer obigen Untersuchung auch beim Lösen der C-Tests zu bestätigen scheint.

7. Datenerhebung und Analyse der Daten

Im Folgenden wird auf die im Rahmen dieser Studie erhobenen Daten eingegangen. In einem ersten Schritt wird das gesamte Testdesign beschrieben, und in einem zweiten Schritt werden die Analysen einzelner Testverfahren aus der Testbatterie dargestellt. Sowohl einzelne Tests als auch Testkombinationen werden auf ihre Testgütekriterien hin überprüft. Anschließend werden Korrelationsberechnungen für die verschiedenen Testkombinationen (C-Tests/Stolperwörtertests/MC-Lesetests) dargestellt.

Die Datenerhebung an der Universität Duisburg-Essen erstreckte sich über einen Zeitraum von drei Semestern. Die Daten wurden im Wintersemester 2010/2011, im Sommersemester 2011 und im Wintersemester 2011/2012 erhoben. Die Erhebung wurde von der Verfasserin der Arbeit in Anwesenheit der Dozierenden in verschiedenen Vorlesungen und Seminaren durchgeführt. Für die Testdurchführung wurden zwischen 45 und 60 Minuten verwendet. Die Durchführungszeit variierte, weil verschiedene Testverfahren kombiniert wurden, die unterschiedliche Bearbeitungszeiten verlangen¹⁰³. Tabelle 13 gibt einen Überblick über die verwendeten Testverfahren.

Tabelle 14: Erhebungen an der Universität Duisburg-Essen

Erhebung	Testkombinationen
WS 2010/2011	4 C-Tests (100 Items) und schulischer Stolperwörtertest (60 Items)
SS 2011	1 modifizierter Stolperwörtertest für Erwachsene (35 Items) und 4 C-Tests (100 Items) <hr/> <i>veränderte Reihenfolge</i> 4 C-Tests (100 Items) und 1 modifizierter Stolperwörtertest für Erwachsene (35 Items)
WS 2011/2012	4 C-Tests (100 Items) , 1 MC-Lesetest (10 Items), 1 MC-Zuordnungsaufgabe (10 Items)

Anmerkung: Die Reihenfolge der vier C-Tests ist bei allen Kombinationen folgende: 1. C-Test (Schulwesen), 2. C-Test (Werther), 3. C-Test (Motivationsformen), 4. C-Test (Frontalunterricht).

¹⁰³ Die verwendeten Tests inkl. Instruktion und Lösungsschablonen sind im Anhang A, Teil 2 zu finden.

Im Folgenden werden die Tests einzeln und in Kombinationen analysiert.

7.1 C-Tests und Lesetests: Gruppen und Daten

Zur Vereinfachung der Auswertung und Analyse der Daten wurden zunächst in einem ersten Schritt die C-Tests der vier Stichproben und in einem zweiten Schritt die damit verwendeten Lesetests ausgewertet. Für die C-Tests wurden Lösungsschablonen genutzt, die dem *Sprachkompetenzprojekt Diagnose & Förderung* entnommen wurden¹⁰⁴.

Auswertung der Items im C-Test:

Jede Lücke im C-Test wurde separat codiert, wobei zwei Codes möglich waren (richtig – 1; falsch – 2). Für jede richtig ausgefüllte Lücke wurde 1 Punkt vergeben. Wurde die Lücke semantisch, grammatisch oder orthografisch falsch rekonstruiert oder gar nicht ausgefüllt, wurden 0 Punkte vergeben.

Für die Auswertung der C-Tests wurde eine Skala mit Rohpunktwerten benutzt, die für jeden einzelnen C-Test eine Spannweite von 0 bis 25 Punkten hat, da jeder C-Test 25 Lücken enthält. Für alle vier C-Tests hat die Skala eine Spannweite von 0 bis 100 Punkten.

Auswertung der Items in den Stolperwörtertests:

Bei den Stolperwörtertests wurden *nur* die Originallösungen als richtige Variante akzeptiert. Jedes einzelne Item (schulischer Stolperwörtertest: 60 Items, modifizierter Stolperwörtertest: 35 Items) wurde ausgewertet und entweder mit Code 1 (richtige Lösungen) oder Code 2 (falsche Lösungen) codiert. Nicht bearbeitete Items wurden mit dem Code 3 codiert (dazu in Kap.7.3.2 und 7.3.3).

Auswertung der Items in den MC-Lesetests:

Die MC-Lesetests wurden nach dem Lösungsschlüssel ausgewertet, der vom TestDaF-Institut freigegeben wurde¹⁰⁵. Die Codierung erfolgte analog zu den

¹⁰⁴ Die verwendeten Lösungsschablonen befinden sich in Anhang A, Teil 2.

¹⁰⁵ Der Lösungsschlüssel befindet sich in Anhang A, Teil 2.

anderen Tests; richtige Antworten wurden mit dem Code 1 und falsche mit dem Code 2 codiert.

Die gesamte statistische Auswertung erfolgte mit dem Datenverarbeitungsprogramm SPSS (Version 18) und den Programmen R und ACER ConQuest (für die Rasch-Analysen).

Da die Gesamtgruppe in Bezug auf die Muttersprache nicht homogen ist und sowohl aus Studierenden mit der Muttersprache Deutsch als auch aus Studierenden mit anderen Muttersprachen besteht, wird bei der Analyse der Daten auf diese zwei Gruppen vergleichend eingegangen.

Der Migrationshintergrund wurde anhand folgender Variablen durch entsprechende Fragen bestimmt: Muttersprache der beiden Elternteile und Muttersprache des Probanden (vgl. dazu ausführlich Kap. 8.2 und Fragebogen im Anhang A, Teil 1)¹⁰⁶. Die Daten der Studierenden mit Migrationshintergrund werden in Bezug zum Gruppennormwert (vgl. Kap. 4.9) und zur Leistung der Studierenden mit der Muttersprache Deutsch interpretiert. Insgesamt wurden 887 Probanden mit verschiedenen Testkombinationen (s. Tabelle 7.2) getestet. Mit dem Fragebogen zur Person konnten 11 bis 16 weitere Hintergrundvariablen bezüglich Schulausbildungsweg, Abiturnoten in verschiedenen Fächern und Selbsteinschätzung der Teilkompetenzen (schriftliche Kompetenz, Lesekompetenz) erfasst werden¹⁰⁷. Mithilfe dieser Hintergrundvariablen werden in Kap. 8 verschiedene Hypothesen überprüft. In Tabelle 14 sind die soziodemografischen Merkmale der verwendeten Stichproben dargestellt.

¹⁰⁶ Ausführliche Daten bzgl. der Sprachbiografien der Probanden (zum Beispiel Kommunikationssprache in der Familie, Geburtsort der Probanden bzw. der Eltern) konnten im Rahmen dieser Untersuchung nicht erhoben werden.

¹⁰⁷ Die Zahl der Fragen variiert, da durch Modifizierung und Erweiterung einiger Fragen im Laufe der Untersuchung verschiedene Fragebogenversionen entstanden sind.

Tabelle 15: Das Testdesign der Studie und die Zusammensetzung der Studierenden in verschiedenen Stichproben

Stichproben (STPB)	Häufigkeit	Geschlecht	Migrationshintergrund (Mh)
STPB1: 4 C-Tests und 1 schulischer Stolperwörtertest	223	m 41 (18,4 %) w 182 (81,6 %)	mit Mh 81 (36,3 %)
STPB2: 4 C-Tests und 1 modifizierter Stolperwörtertest, Reihenfolge (C- Tests/Stolperwörtertest);	246	m 65 (26,4 %) w 181 (73,6 %)	mit Mh 103 (41,9 %)
STPB3: 4 C-Tests und 1 modifizierter Stolperwörtertest Reihenfolge (Stolperwörtertest/C-Tests)	245	m 75 (30,6 %) w 169 (69 %)	mit Mh 76 (31 %)
STPB4: 4 C-Tests und TestDaF-MC-Lesetests (MC-Test/MC- Zuordnungsaufgabe)	173	m 38 (22 %) w 134 (77,5)	mit Mh 73 (42,2 %)
Gesamt	887	m 219 (24,7 %) w 666 (75,1 %) k. A. 2 (0,2 %)	mit Mh 333 (37,5 %) k.A. 8 (0,9 %)

7.2 C-Tests: Item- und Reliabilitätsanalysen

Die eingesetzten C-Tests werden im Weiteren auf ihre Schwierigkeit, Trennschärfe, Interkorrelationen und Reliabilität hin untersucht. Ein wesentliches Ziel der Analysen ist es, die Testgüte in Bezug auf Reliabilität und Validität zu beschreiben und zu erhöhen. Die Validität der verwendeten C-Tests soll auf dem korrelativen Weg mithilfe ausgewählter Lesetests überprüft werden, die in Kap. 5.4.1 und 5.4.2 dargestellt wurden. Für die statistischen Berechnungen wurden die Daten von 887 Studierenden verwendet.

Analysiert man die erhobenen C-Test-Daten auf die durchschnittlichen Leistungen in den jeweiligen C-Tests (innerhalb der C-Testbatterie) und ihre Schwierigkeiten hin, so ergibt sich folgendes Bild (s. Tabelle 15):

Tabelle 16: Mittelwerte, Schwierigkeiten und Standardabweichungen von C-Tests

C-Tests	Mittelwert	Schwierigkeit (p)	Standardabweichung (s)
C-Test1	14,5	58,0 (T4)	5,678
C-Test2	15,9	63,5 (T1)	6,136
C-Test3	15,2	60,9 (T2)	6,269
C-Test4	14,3	57,2 (T3)	4,801
Gesamt	59,90	59,9	20,1

In Tabelle 15 werden zusätzlich zu den berechneten Mittelwerten Schwierigkeiten und Standardabweichungen der einzelnen C-Tests angegeben. In der dritten Spalte werden die Schwierigkeitsindizes (p) der einzelnen C-Tests und der gesamten C-Testbatterie dargestellt. Der Schwierigkeitsindex (p) gibt den prozentualen Anteil der Testitems an, den ein Proband durchschnittlich lösen konnte. Ein niedriger Schwierigkeitsgrad charakterisiert eine schwierige Aufgabe (einen schwierigen C-Test), ein hoher eine leichte bzw. einen leichten C-Test (vgl. Wockenfuß, 2009, S. 156). Die dargestellten Werte zeigen, dass die C-Tests für die untersuchte Gruppe mittelschwer sind.

Die Schwierigkeiten (p) für die einzelnen C-Tests (Teilttests) wurden mit folgender Formel berechnet¹⁰⁸:

$$p_j = \frac{\bar{x}}{n_j} \times 100$$

Dabei bedeutet: **p_j**: Schwierigkeit der einzelnen C-Tests; **x_j**: arithmetisches Mittel der Rohwerte des C-Tests **j**; **n_j**: Anzahl der Lücken im einzelnen C-Test.

Die Schwierigkeitsgrade der Texte (hier C-Tests) sollten laut Angaben von ALTE¹⁰⁹ zwischen 0.33 und 0.67 oder zwischen 0.20 und 0.80 liegen. Die Bestimmung der Schwierigkeit der einzusetzenden Aufgaben bzw. Tests richtet sich nach dem

¹⁰⁸ Vgl. dazu Raatz & Klein-Braley, 1985, S. 21; Wockenfuß, 2009, S. 156.

¹⁰⁹ ALTE-Handreichungen für Testautoren, Modul 2, S. 16f.

Zweck, zu dem der Test eingesetzt werden soll. In der vorliegenden Studie sollen die Aufgaben weder zu leicht noch zu schwierig sein, was bedeutet, dass Tests mittlerer Schwierigkeit zwischen kompetenten und weniger kompetenten Probanden besser differenzieren können.

Die ermittelten Schwierigkeitsindizes der vier C-Tests variieren in engen Grenzen im mittleren Bereich zwischen 0.57 und 0.63. In Tabelle 7.3 sind die einzelnen C-Tests so angeordnet, wie sie in der Testbatterie im Sprachkompetenzprojekt verwendet wurden (vgl. die Nummerierung in der ersten Spalte der Tabelle). In der dritten Spalte befinden sich die Werte, die für die vorliegenden Daten berechnet wurden. In der Textanordnung sind einige Abweichungen festzustellen.

Der C-Test Nr. 2 scheint für die getestete Gruppe der leichteste zu sein, und der C-Test Nr. 1 würde an dritter Stelle kommen. In der Spalte *Schwierigkeit* wurde in Klammern eine neue Textanordnung markiert. Da in der vorliegenden Untersuchung die C-Tests nicht vorgetestet wurden, um eventuelle Abweichungen in den Schwierigkeiten zu untersuchen, wurde die Reihenfolge der C-Tests aus dem Sprachkompetenzprojekt übernommen, so wie sie in der ersten Spalte markiert ist.

Im Folgenden soll überprüft werden, wie hoch einzelne C-Tests miteinander korrelieren. Die Interkorrelation der einzelnen C-Tests, über die in Tabelle 16 berichtet wird, gibt an, ob die einzelnen C-Tests gleiche Eigenschaften erfassen und wie hoch die Übereinstimmung bei ihrer Messung ist. Beim C-Test müssen nach Raatz & Klein-Braley (1994) die einzelnen C-Tests relativ hoch miteinander korrelieren, da er ein homogener Test ist und inhaltlich ähnliche Merkmale erfasst (Wockenfuß, 2009, S. 158).

Tabelle 17: Interkorrelationen der einzelnen C-Tests

Einzelne C-Tests	T2	T3	T4	r (it), corr
T1	,68	,71	,68	,77
T2		,70	,67	,76
T3			,75	,80
T4				,79

r (it), corr – korrigierte Korrelationen

In der letzten Spalte der Tabelle 16 werden die berechneten Trennschärfekoeffizienten dargestellt. Der Trennschärfekoeffizient eines C-Tests kennzeichnet den Zusammenhang zwischen der erbrachten Leistung im jeweiligen C-Test und dem Gesamtscore, der sich aus den Ergebnissen von vier C-Tests summiert (vgl. Wockenfuß, 2009, S. 158). Ein hoher Trennschärfekoeffizient bedeutet, dass der C-Test gut zwischen leistungsstarken und leistungsschwachen Probanden unterscheidet; das heißt, der Wert nähert sich dem Wert von + 1 (dazu ALTE, Modul 2, S. 16ff.). Zur Berechnung der Trennschärfe wird die Korrelation zwischen den Antworten eines C-Tests und dem Gesamtwert (summiert aus vier C-Tests) berechnet. Um die Selbstkorrelation des jeweiligen C-Tests (z. B. C-Test 1) mit dem Gesamtwert herauszurechnen, wird die ‚part-whole‘ (*rjt, corr*) berechnet. Das bedeutet, dass der Wert aus dem jeweiligen C-Test (z. B. C-Test 1) mit dem um diesen Wert reduzierten Gesamtwert (Gesamtwert minus Wert aus dem C-Test 1) korreliert wird. Dieser Wert zeigt an, ob dieser C-Test in der Gesamtbatterie zwischen der Leistung der Probanden gut trennen kann. Vorausgesetzt wird, dass jeder C-Test das gleiche Merkmal bzw. die gleichen Merkmale misst.

Bei der Beschreibung und Interpretation der dargestellten Werte muss berücksichtigt werden, dass drei Itemkennwerte (Testschwierigkeiten, Korrelationen innerhalb der Testbatterie und die korrigierte Korrelation) nicht unabhängig voneinander sind. Wie aus Tabelle 16 ersichtlich, sind die Interkorrelationskoeffizienten hoch; die Werte variieren zwischen 0.68 und 0.75. Das spricht dafür, dass die einzelnen C-Tests weitgehend ähnliche Eigenschaften messen (z. B. Sprachkompetenz). In der dargestellten C-Testbatterie haben der

dritte und vierte C-Test die beste Trennschärfe. Die Abweichungen in der berechneten Trennschärfe bei den anderen C-Tests sind sehr gering und fast identisch.

Aufgrund der ermittelten Kennwerte, die durchaus als befriedigend zu interpretieren sind, kann festgehalten werden, dass die verwendeten C-Tests eine globale Differenzierung innerhalb der Probandengruppe ermöglichen.

Bei der Berechnung der Reliabilität für die verwendeten vier C-Tests erreicht Cronbachs Alpha den Wert von $\alpha = 0.897$ und für standardisierte Items $\alpha = 0.901$, was für die verwendete Testbatterie und den Untersuchungszweck als zufriedenstellend interpretiert wird (vgl. Bortz, 2005).

Festlegung der Leistungsbereiche:

Um die Leistungen der Studierenden verschiedenen Leistungsgruppen zuzuordnen, wurden die Daten normiert und anschließend skaliert. Die Normierungsbereiche wurden aus dem *Sprachkompetenzprojekt Diagnose & Förderung* übernommen, da dieser Untersuchung eine ähnliche Zielgruppe zur Verfügung steht wie dem Sprachkompetenzprojekt. Für die Normierung der Daten wurden die Daten von ca. 2.000 Lehramtsstudierenden genutzt, was bedeutet, dass sowohl der ermittelte Referenznormwert als auch die festgelegten Normbereiche statistisch stabil sind (vgl. dazu auch die Beschreibungen in Kap. 4.9). Allerdings können die Abstände zwischen den Leistungsbereichen nicht interpretiert werden, d. h. erreicht Person A den C-Testwert, der der höchsten Leistungsstufe zuzuordnen ist und Person B einen Wert, der um 2 Punkte niedriger ist und somit einer niedrigeren Stufe zugeordnet wird, kann nicht darauf geschlossen werden, dass Person A eine bessere Sprachkompetenz hat als Person B. Die Leistungsstufen dienen nur der statistischen Rangordnung innerhalb der getesteten Gruppe bzw. der Gesamtgruppe der Lehramtsstudierenden im Grundstudium. Anhand der Leistungsstufen kann geschätzt werden, ob die erzielte Leistung über der durchschnittlichen Leistung der Studierenden liegt und um wie viele Standardabweichungen die erbrachte Leistung vom ermittelten Referenznormwert entfernt ist. Weicht die Leistung um zwei

Standardabweichungen nach unten ab, wird dem Studenten ein großer sprachlicher Förderbedarf unterstellt (vgl. dazu Publikationen von Scholten-Akoun & Baur, 2012; Scholten-Akoun, Kuhnen & Mashkovskaya (2012); Scholten-Akoun, Tischmeyer & Mashkovskaya, 2013).

Die Normierung der C-Testdaten basiert auf den IRT-Analysen. Umfassende Untersuchungen zur Anwendung von IRT-Modelle für die C-Tests sind bei Eckes (2006, 2010b, 2011a, b), Eckes & Grotjahn (2006a, b) und Grotjahn (2007) beschrieben. Umfangreiche und sehr detaillierte Analysen in Bezug auf die Anwendung der IRT-Modelle für die in der vorliegenden Studie verwendeten C-Tests für Studierende finden sich bei Baur, Mashkovskaya & Weirich in der Projektpublikation von Bremerich-Vos et al. (2013).

Da die Normierung und Skalierung der C-Tests für die gleiche Gruppe der Studierenden aus dem Projekt entnommen wurde, wird auf die Skalierung der C-Test-Daten nicht näher eingegangen, da die Daten ähnlich sind.

In Tabelle 30 sind die Leistungsstufen dargestellt, die sich anhand des ermittelten Referenznormwertes und der Standardabweichungen definieren lassen. Die mithilfe des Auswertungsmodells (Ratingskalenmodell)¹¹⁰ berechneten Normwerte werden in Rohwerte (C-Testpunkte) umgewandelt, mit denen die Leistung der Studierenden verglichen werden kann.

Für das Normierungssample ergeben sich ein Mittelwert von -0,008 und eine Standardabweichung von 0,923. Damit lassen sich z. B. folgende Normbereiche definieren:

Tabelle 18: Festlegung der Daten für die C-Tests für Lehramtsstudierende

Stufe	Bereich	Normwert	Rohwert
1	>2 Standardabweichungen unter dem Normwert	>1,8386	Personen mit einem Rohwert > 89
2	1 bis 2 Standardabweichungen unter dem Normwert	0,9153 bis 1,8386	Personen mit einem Rohwert von 81 – 89
3	±1 Standardabweichung um den Normwert	-0,9315 bis 0,9153	Personen mit einem Rohwert von 54 – 80
4	1 bis 2 Standardabweichungen über dem Normwert	-1,855 bis -0,9315	Personen mit einem Rohwert von 39 – 53
5	>2 Standardabweichungen über dem Normwert	< -1,855	Personen mit einem Rohwert von 0 – 38

¹¹⁰ Das Ratingskalenmodell wird der Items-Response-Theory (IRT) zugeordnet.

7.3 Vergleich zwischen den C-Tests und Lesetests: Reliabilitätsanalysen und Korrelationsberechnungen

Da C-Tests in der vorliegenden Untersuchung zusammen mit verschiedenen Lesetests verwendet wurden, bilden diese Kombinationen im Sinne der Testtheorie ein gemeinsames Konstrukt bzw. gemeinsame Konstrukte, die aus mehreren Items (Tests) bestehen. In welchem Zusammenhang diese Items zueinander stehen und ob die Konstrukte zuverlässig messen, wird mithilfe von Reliabilitätsanalysen bestimmt. Die Reihenfolge der Analysen wird so beibehalten, wie sie in der Tabelle 13 in Kap. 7.1 präsentiert wurde.

7.3.1 C-Tests und LT1 (schulischer Stolperwörtertest)

***Hypothese 4:** Bei der kombinierten Verwendung von C-Tests und des schulischen Stolperwörtertests bestehen Korrelationen.*

Die Grundlage für diese Testkombination bilden vier C-Tests und ein schulischer Stolperwörtertest, der hier zur besseren Orientierung als LT1 bezeichnet wird. Der schulische Stolperwörtertest besteht aus 60 kurzen Sätzen (Items), in denen Fehler verschiedener Art (syntaktisch, grammatisch, orthografisch) eingebaut sind (ausführliche Informationen zum Konstrukt in Kap. 5.4.2). Die Studierenden sollten unter zeitlichem Druck die Fehler identifizieren, indem sie die entsprechende Stelle im Satz markieren (z. B. das falsch verwendete Wort durchstreichen). Der komplette Test inkl. Lösungsschlüssel findet sich im Anhang A, Teil 2¹¹¹. In einem ersten Schritt wurden den Studierenden vier C-Tests und erst in einem zweiten Schritt der Stolperwörtertest vorgelegt. Für den Stolperwörtertest standen den Probanden 5 Minuten, für vier C-Tests 20 Minuten zur Verfügung. Bei der Auswertung der Ergebnisse wurde, wie bereits beschrieben, für jeden richtig bearbeiteten Satz ein Punkt vergeben, für jeden

¹¹¹ Bei der Auswertung des Stolperwörtertests wurden Lösungsschablonen verwendet, die im Internet von W. Metze zur Verfügung gestellt wurden;
<http://www.wilfriedmetze.de/html/stolper.html>, zuletzt gesehen am 13.05.2013.

falsch oder nicht bearbeiteten kein Punkt. Bei der Analyse der Daten ergaben sich folgende Auswertungsmodi¹¹²:

Modus 1: *Nicht bearbeitete Items (hier: Sätze) am Ende des Lesetests werden als Missings codiert, d. h. sie werden nicht als Fehler gewertet.*

Modus 2: *Nicht bearbeitete Items am Ende des Lesetests werden als Falschantworten codiert.*

Modus 1: Bei der Anwendung des ersten Auswertungsmodus stellte sich heraus, dass alle Items (60 Sätze) für die Probanden zu leicht sind, da fast alle Sätze von den Studierenden zu 100 % korrekt gelöst wurden. Für die Analyse der Daten wurden vorab 10 Items ausgeschlossen; 13 Items haben schlechte Fitkriterien (Infit-Werte)¹¹³. Insgesamt wurden 37 Items analysiert. Die berechnete Reliabilität beträgt Cronbachs Alpha $\alpha = 0.471$ und ist als nicht zufriedenstellend zu interpretieren (vgl. Analysen dazu im Anhang A, Teil 3).

Modus 2: Bei der Anwendung des zweiten Auswertungsmodus¹ wurden unbearbeitete Sätze am Ende des Tests als Falschantworten gewertet. Geht man davon aus, dass kompetente Probanden den Test schnell und korrekt lösen, fiel es den weniger kompetenten Lesern schwer, die Sätze am Ende der Tests zu bearbeiten, da sie möglicherweise langsamer lesen. Wertet man diese Items als Falschantworten, werden langsame Leser dadurch benachteiligt (vgl. Analysen dazu im Anhang A, Teil 3).

Bei der Anwendung des zweiten Modus¹ ergab sich, dass die Items an hinteren Positionen im Test (wie erwartet) schwerer sind, da nicht jeder Proband gleich schnell liest und diese Items bearbeitet.

¹¹² Beide Auswertungsmöglichkeiten werden nur hinsichtlich ihrer empirischen Ergebnisse verglichen. Die Auswertungsmöglichkeiten implizieren aber inhaltlich verschiedene Konstruktdefinitionen. Fehlende Antworten am Testende als falsch zu bewerten bedeutet zum Beispiel, dass die Bearbeitungsgeschwindigkeit als Komponente der Testleistung mitbewertet wird. Für den Fall, dass die Antworten, die nicht bearbeitet wurden, herausgerechnet werden, wird die Geschwindigkeit nicht mitbewertet.

¹¹³ Infit-Werte sind ein Gütemaß der IRT-Modelle.

Tabelle 19: Korrelationen zwischen den C-Tests und dem Stolperwörtertest (LT1)

Verwendete Tests	Stolperwörtertest Auswertungsmodus 1	Stolperwörtertest Auswertungsmodus 2
vier C-Tests	$r = 0.45$	$r = 0.52$

Die Korrelationen zwischen den C-Tests und den Stolperwörtertests variieren je nach Auswertungsmodus. Obwohl der Stolperwörtertest für die Untersuchungsgruppe zu leicht ist, bestehen niedrige bis mittlere Korrelationen. Dieser Befund lässt annehmen, dass der Stolperwörtertest und der C-Test bestimmte Bereiche der Teilkompetenz auf ähnliche Weise messen (vgl. Kap. 6.4.2). Auffallend ist jedoch, dass sie trotz der unzureichenden Schwierigkeit für die Testgruppe und der zu geringen Reliabilität beider Testverfahren (C-Tests und Stolperwörtertest) schwach bis mittelstark miteinander korrelieren.

Zusammenfassend lässt sich festhalten, dass der schulische Stolperwörtertest in dieser Form trotz der festgestellten mittleren Korrelationen für die Untersuchungsgruppe nicht geeignet ist, da die Mehrheit aller Items zu einfach ist und die Messgenauigkeit innerhalb der Gruppe dadurch eingeschränkt ist (vgl. dazu auch Kap. 5.4.1). Der Test bildet das Konstrukt Leseverständnis, Lesegenauigkeit und Lesegeschwindigkeit nur in unzureichender Weise und mit zu geringer Reliabilität ab.

Vor diesem Hintergrund wurde entschieden, einen modifizierten Stolperwörtertest für Erwachsene zu entwickeln, der sowohl inhaltlich als auch sprachlich dem Niveau der Studierenden angepasst ist. Ob der modifizierte C-Test mit den C-Tests besser korreliert, wird in Kap. 7.3.2 überprüft.

Die Hypothese wurde bestätigt, zwischen den C-Tests und dem schulischen Stolperwörtertest bestehen schwache bis mittlere Korrelationen.

7.3.2 C-Tests und LT2 (modifizierter Stolperwörtertest für Erwachsene)

***Hypothese 5:** Bei der kombinierten Verwendung der C-Tests und des modifizierten Stolperwörtertests für Erwachsene bestehen Korrelationen.*

Angenommen wird, dass der Zusammenhang zwischen den C-Tests und dem modifizierten Stolperwörtertest höher ausfällt als beim Einsatz des schulischen Stolperwörtertests.

Der modifizierte Stolperwörtertest für Erwachsene wird im Folgenden LT2 genannt. Dieser Test wurde für die vorliegende Studie entwickelt, um zu überprüfen, ob sich durch eine inhaltliche und sprachliche Modifikation des Stolperwörtertests die Lesekompetenz der Studierenden besser abbilden lässt. Angenommen wird, dass sich die obengenannten Modifikationen zum einen positiv auf die Reliabilität auswirken und zum anderen auf die Validität des Tests, die durch die Korrelationen überprüft wird.

Der Stolperwörtertest für Studierende wurde, wie bereits in Kap. 6.4.2 beschrieben, auf 35 Items verkürzt¹¹⁴. Er besteht aus Sätzen, die aus zwei pädagogischen Artikeln stammen. Ähnliche Texte wurden auch für die Entwicklung der C-Tests für Lehramtsstudierende verwendet. Die Fehlerkategorien, wie sie im schulischen Stolperwörtertest vorkommen, wurden auch in den Stolperwörtertest für Studierende eingebaut. Ein Beispiel dieses Tests befindet sich im Anhang A, Teil 2.3. Den Studierenden wurden 5 Minuten Zeit für die Bearbeitung gegeben.

Um Zusammenhänge zu modellieren und zu berechnen, werden der Stolperwörterlesetest und vier C-Tests in einem zweidimensionalen Modell abgebildet.

Als Messmodell für die C-Tests wird ein Ratingskalenmodell zugrunde gelegt und für die Stolperwörtertests ein Raschmodell, da in der Literatur zu den C-Tests die Verwendung von Raschmodellen aufgrund der lokalen Abhängigkeiten zwischen den Items (einzelnen C-Test-Lücken) kritisiert wird¹¹⁵ (vgl. Eckes, 2010b). Beide

¹¹⁴ Wie es auch in der STOLLE-Untersuchung praktiziert wird (vgl. dazu Kap. 6.5.2).

¹¹⁵ Vgl. dazu die Analyse der Q3-Statistik in Kap. 7.4.2.

Modelle können für dichotome Items (für die Auswertungsmöglichkeiten richtig vs. falsch) verwendet werden. In einem ersten Schritt werden latente Korrelationen zwischen dem nach Modus 1 ausgewerteten Stolperwörtertest und den vier C-Tests berechnet und in einem zweiten Schritt Korrelationen zwischen dem nach dem Modus 2 ausgewerteten Stolperwörtertest und den vier C-Tests.

Modus 1: Die Korrelationen auf latenter Ebene¹¹⁶ zwischen den vier C-Tests und dem modifizierten Stolperwörtertest für Studierende betragen $r = 0.802$ und sind als hoch zu interpretieren. Die gesamte Testbatterie wurde als polytome Items codiert und Items im Stolperwörtertest als dichotome Items¹¹⁷.

Modus 2: Die Korrelationen auf latenter Ebene zwischen den vier C-Tests und dem modifizierten Stolperwörtertest für Studierende betragen $r = 0.718$. Die Codierung der Items erfolgt ähnlich wie im Modus 1¹¹⁸.

Auffallend ist, dass die Auswertung nach verschiedenen Modi auch unterschiedliche Korrelationswerte produziert. Der Zusammenhang zwischen beiden Testinstrumenten sinkt, sobald die Speedkomponente in die Bewertung eingebaut wird. Der Stolperwörtertest misst demzufolge nicht mehr nur das Lesen an sich, sondern auch Lese- oder Verarbeitungsgeschwindigkeit, also ein neues, gemeinsames Konstrukt. Dieses neue Konstrukt unterscheidet sich von dem Lesekonstrukt, das der C-Test misst, stärker (vgl. die Korrelationswerte), eben durch diese zusätzliche Speedkomponente.

Sinkt der Zusammenhang zwischen beiden Tests, wie auch im Modus 2 der Fall, könnte das gemeinsame Konstrukt, das beide Tests (C-Tests und Stolperwörtertests) messen, nicht nur die Teilkompetenz Lesen umfassen, sondern auch andere Teilkompetenzen, die sich z. B. in dieser Form durch die C-Tests oder den Stolperwörtertest allein nicht auf ähnliche Weise erfassen lassen. Ähnlich wie in Kap. 7.3.1 werden für die Analyse der Daten zwei Auswertungsmodi verwendet:

¹¹⁶ Das andere sind Korrelationen auf manifester Ebene. Diese können jedoch aufgrund von Messfehlern überschätzt (i.d.R. unterschätzt) werden.

¹¹⁷ In diesem Fall der Auswertung von C-Tests erhält man statt 100 dichotom codierter Items (25 Items je C-Test) 4 polytom codierte Items mit jeweils 25 möglichen diskreten Werten.

¹¹⁸ Für die Analyse der C-Tests wird ein Ratingskalenmodell verwendet.

Modus 1: *nicht bearbeitete Items am Ende des Lesetests wurden als Missings codiert*

Modus 2: *nicht bearbeitete Items am Ende des Lesetests wurden als Falschantworten codiert*

Für die Analyse des ersten Modus wurde, wie bereits erwähnt, das Raschmodell aus der Familie der probabilistischen Testmodelle oder Item-Response-Theorie (IRT) verwendet. Vorteil der IRT ist, dass sich verschiedene Konstrukte (z. B. Lesekompetenz) auf latenter Ebene modellieren und sich die damit verbundenen Annahmen explizit testen lassen. Lesekompetenz als ein latentes Konstrukt zu definieren impliziert die Annahme, dass es nicht direkt zu beobachten bzw. zu messen ist. Eine Messung kann daher nur auf indirektem Wege über eine Modellvorstellung erfolgen. Die Modellvorstellung hier lässt sich folgendermaßen formulieren: Die Lesekompetenz einer Person beeinflusst ihr Verhalten bzw. ihr Antwortverhalten auf die Items eines Lesetests. Je höher die unterstellte Lesekompetenz einer Person, umso höher ist die Wahrscheinlichkeit, dass sie ein Item des Tests richtig löst. Diese sehr einfache Modellvorstellung erlaubt es nun, zu prüfen, inwieweit die Items eines Tests mit ihr in Übereinklang stehen. Items, für die dies nicht der Fall ist, sind ungeeignet, das entsprechende Konstrukt reliabel zu messen. Dies kann an zwei Beispielen des Stolperwörtertests veranschaulicht werden.

Ein Item aus 35 vorhandenen Items wurde aus der Analyse ausgeschlossen, weil es von allen Probanden zu 100 % gelöst wurde und innerhalb der Gruppe nicht differenzierte. Ein solches Item, das *von allen* Probanden gelöst wird, kann nicht zwischen fähigen und weniger fähigen Personen diskriminieren. Es misst offenbar keine Fähigkeit. Ein weiteres Item hat einen auffälligen Infit-Wert. Dies kann auch für Items mittlerer Lösungshäufigkeit der Fall sein. Wenn etwa ein Item von fähigen und weniger fähigen Personen mit nahezu der gleichen Wahrscheinlichkeit gelöst werden kann, widerspricht es der Annahme des Raschmodells und differenziert ebenfalls nicht zwischen fähigen und weniger fähigen Personen. Der Grad dieses ‚Modell-Misfits‘ spiegelt sich im sogenannten Infit wider, der für jedes Item einzeln berechnet werden kann. Wünschenswert ist dabei, dass dieser Koeffizient den Wert von 1,15 nicht übersteigt. Da der

Infitwert mit 1.25 nur geringfügig über dem Grenzwert lag, wurde das Item trotzdem in der Analyse beibehalten. Die berechnete Reliabilität beträgt Cronbachs Alpha $\alpha = 0.76$ und ist als zufriedenstellend zu interpretieren.

Ein weiteres Indiz für die Güte des Modells ist die Q3-Statistik. Wie oben beschrieben, soll die Lösungshäufigkeit im Raschmodell *ausschließlich* von der Schwierigkeit des Items und der Fähigkeit der Person abhängen. Zumindest für den C-Test wäre diese Voraussetzung kritisch zu hinterfragen. Da der inhaltliche Kontext des im C-Test vorgelegten Textes umso verständlicher wird, je mehr Items eine Person bereits rekonstruiert hat und da ein besseres Verständnis dieses inhaltlichen Kontextes sich positiv auf die Lösungswahrscheinlichkeit weiterer Lücken auswirken könnte, kann argumentiert werden, dass die Wahrscheinlichkeit, im Test eher an hinterer Position stehende Items (Lücken) zu lösen, *zusätzlich* davon abhängt, wie viele Items bereits gelöst worden sind. Wenn das der Fall wäre, würde die Modellannahme ebenfalls verletzt sein. Um dies zu prüfen, kann die Q3-Statistik verwendet werden, welche die Residualkorrelationen aller Items abbildet. Das Prinzip dieser Prüfstatistik lässt sich folgendermaßen zusammenfassen: Es wird zunächst angenommen, dass das Raschmodell gilt, d.h. die Antwortwahrscheinlichkeit jedes Items hängt nur von der Schwierigkeit des Items und der Fähigkeit der Person ab. Wenn also Items miteinander korrelieren, dann nur deshalb, weil die Personen unterschiedlich fähig sind. Hätten alle Personen, die den Test bearbeiten, dieselbe latente Fähigkeit, sollten die Items untereinander nicht korrelieren. Die Statistik prüft nun die Korrelation unter statistischer Konstanthaltung der Personenfähigkeit. In der Literatur werden dabei verschiedene Grenzwerte genannt, die als kritisch zu interpretieren sind. Residualkorrelationen über 0.25 gelten meist als auffällig, Korrelationen von über 0.35 werden als kritisch interpretiert. Im vorliegenden Fall liegen ca. 10 % der Fälle über der auffälligen Grenze.

Bei der Anwendung des zweiten Auswertungsmodus wurden unbearbeitete Sätze am Ende des Tests als Falschantworten gewertet. Die Items an hinteren Positionen sind, wie erwartet, schwerer, da nicht alle Probanden die Items am Ende des Tests bearbeiten. Der Item-Fit erreicht Werte bis 1.7, was für die Anpassung des Modells als schlecht zu interpretieren ist.

Die Anwendung des zweiten Auswertungsmodells erscheint aus zwei Gründen ungeeignet: Geht man davon aus, dass kompetente Probanden den Test schnell und korrekt lösen, fiel es weniger kompetenten Lesern schwerer, die Sätze am Ende der Tests zu bearbeiten, da sie möglicherweise langsamer lesen. Wertet man diese Items als Falschantworten, werden langsame Leser benachteiligt. Im Folgenden werden die berechneten Korrelationen dargestellt.

Tabelle 20: Korrelationen zwischen den C-Tests und dem Stolperwörtertest (LT2)

verwendete Tests	Stolperwörtertest	Stolperwörtertest
	Auswertungsmodus 1	Auswertungsmodus 2
vier C-Tests	$r = 0.82$	$r = 0.718$

Aus Tabelle 20 wird ersichtlich, dass zwischen den C-Tests und dem modifizierten Stolperwörtertest starke Zusammenhänge bestehen. In Bezug auf die Auswertungsmodi bedeutet dies, dass in den Lesetests nicht nur die Leseleistung, sondern auch die Verarbeitungsgeschwindigkeit überprüft wird. Rechnet man die Speedkomponente heraus, so sinkt auch der Zusammenhang zwischen beiden Tests. Für weitere Analysen wird daher das erste Auswertungsmodell präferiert, da es die Daten offenbar besser beschreibt.

In Bezug auf die Prozessebenen, die beide Tests (C-Tests und Stolperwörtertest) beschreiben, bedeutet der hohe Zusammenhang, dass sich durch den Einsatz der oben beschriebenen Testkombination ähnliche Prozessebenen erfassen lassen. Wie bereits in Kap. 6.2 beschrieben, kann der C-Test auf unterschiedlichen Prozessebenen Fähigkeiten der Studierenden messen. Misst er auf der Wort- und Satzebene, die den hierarchieniedrigeren Prozessebenen zugeordnet sind, so erfasst diese Ebenen auch der Stolperwörtertest. In beiden Tests (vgl. dazu Kap. 5.4.2) kann überprüft werden, ob der Proband auf syntaktischer, semantischer, orthografischer und grammatischer Ebene sein Wissen anwenden kann. Scheitert der Proband bei Identifikation eines semantischen Fehlers im Stolperwörtertest, indem er das falsche Wort streicht, hat er vermutlich auch

beim C-Test bei der Rekonstruktion mancher Lücken Schwierigkeiten. Der hohe Zusammenhang bedeutet, dass hohe Testergebnisse, die im Stolperwörtertest erzielt wurden, auch höheren C-Testleistungen zugeordnet werden. Leistungsdifferenzen (z. B. schwache C-Testleistung vs. hohe Leistung im Stolperwörtertest oder hohe C-Testleistung vs. niedrige Leistung im Stolperwörtertest) kommen nicht vor.

Die Hypothese wurde bestätigt, zwischen den C-Tests und dem modifizierten Stolperwörtertest besteht ein hoher Zusammenhang.

7.3.3 C-Tests und LT3 (Stolperwörtertest für Erwachsene mit veränderter Reihenfolge)

***Hypothese 6:** Die Reihenfolge der Bearbeitung von Tests (C-Test und Stolperwörtertest) hat keine Auswirkungen auf die Testergebnisse.*

Die Grundlage für die Kombination LT3 bilden die gleichen Tests, wie sie in Kap. 7.3.2 beschrieben sind. Der einzige Unterschied, der hier wichtig ist, ist die Reihenfolge, in der die Tests den Studierenden zur Bearbeitung vorgelegt werden. In der vorliegenden Variante wurden in einem ersten Schritt der Stolperwörtertest für Studierende und erst in einem zweiten Schritt vier C-Tests eingesetzt. Die Bearbeitungszeit wurde nicht verändert, sondern ähnlich wie beim LT2 beibehalten. Für die Bearbeitung des Stolperwörtertests standen den Studierenden 5 Minuten und für die Bearbeitung von vier C-Tests 20 Minuten zur Verfügung. Für die Analysen wurden, ähnlich wie in Kap. 7.3.1 und 7.3.2, zwei Auswertungsmodi verwendet.

Modus 1: Bei der Anwendung des ersten Auswertungsmodells wurden unbearbeitete Sätze am Ende des Tests als Falschantworten gewertet. Die Items an hinteren Positionen sind, wie bereits berichtet, schwerer, da nicht alle Probanden Items am Ende des Tests bearbeiten. Die Reliabilität für die vorliegende Kombination mit der veränderten Reihenfolge der Tests sinkt für beide Testteile (C-Test und Stolperwörtertest). Beim C-Test sinkt sie von $\alpha = 0.87$ auf 0.79 und beim Stolperwörtertest von 0.82 auf 0.70.

Bei der Berechnung der Korrelationen zwischen beiden Tests (Stolperwörtertest und C-Tests) ist ebenso ein starker Abstieg festzustellen. Der Unterschied zwischen den berechneten Korrelationen, wenn die C-Tests nach dem Stolperwörtertest vorgelegt werden, beträgt $r = 0.802$. Verändert man die Reihenfolge und lässt die Studierenden erst die C-Tests bearbeiten, dann verändert sich die Höhe des Zusammenhangs zwischen beiden Tests und beträgt $r = 0.524$. Im Folgenden wird auf mögliche Gründe für diese Unterschiede eingegangen.

Modus 2: Analysiert man die Ergebnisse mit dem zweiten Modus, so ergeben sich folgende Unterschiede: Die Reliabilitätswerte sinken für beide Testteile. Bei den C-Tests ist ein kleiner Abstieg von 0.877 auf 0.803 festzustellen. Beim Stolperwörtertest ist der Abstieg ebenso gering: von 0.934 auf 0.931. Beide Reliabilitätswerte sind hoch und werden als zufriedenstellend interpretiert.

Tabelle 21: Korrelationen zwischen den C-Tests und Stolperwörtertest (LT3)

Verwendete Tests	Stolperwörtertest Auswertungsmodus 1	Stolperwörtertest Auswertungsmodus 2
vier C-Tests	$r = 0.524$	0.633

Die in Tabelle 21 dargestellten Korrelationen sind als mittelstark zu interpretieren. Allerdings sinkt, wie oben beschrieben, der Zusammenhang, sobald die Reihenfolge der Tests verändert wird (vgl. Tabelle 7.7 in Kap. 7.3.2). Die Korrelation verringert sich ebenso, wenn die Items an hinteren Positionen im Test als Falschantworten gewertet werden (Auswertungsmodus 1).

Einer der Gründe für die Veränderung der Höhe der Korrelation könnte die Position des Tests in der Testbatterie sein. Angenommen wird, dass Tests, die den Studierenden am Anfang der Testung vorgelegt werden (in unserem Fall vier C-Tests), besser (zum Beispiel: konzentrierter und ordentlicher) bearbeitet werden als Tests, die am Ende zum Einsatz kommen. In der vorliegenden Studie wurde der Stolperwörtertest erst nach der zwanzigminütigen Bearbeitung von C-Tests vorgelegt. Warum derartige Differenzen zustande kommen, ist unklar. Ein Nachlassen der Aufmerksamkeit und Konzentration kann solche gravierenden

Differenzen in der Korrelation (von 0.802 auf 0.524) offenbar nicht erklären, zumal die Testungszeit von 30 Minuten bei erwachsenen Personen zu kurz ist, um Aufmerksamkeitsdefizite in messbarer Form zu erwarten¹¹⁹.

Eine Erklärung für die schlechteren Werte könnte die Motivation der Studierenden sein. Die Durchführung innerhalb der Veranstaltung erfolgte freiwillig. Die Testung wurde am Ende der Veranstaltung, in der letzten halben Stunde, durchgeführt; das könnte die Motivation der Studierenden senken und Auswirkungen auf die Bearbeitung haben. Eine weitere wichtige Rolle bei der Testung könnten der Dozent und/oder der Testdurchführende spielen. Die vorliegende Testung wurde in Anwesenheit der Dozierenden durchgeführt. Bei der Testdurchführung sind einige Dozenten darum bemüht, gute Testbedingungen zu schaffen, andere wiederum sind nicht daran interessiert, dass Studierende kooperieren. Entscheidend ist das für die Testbedingungen und die Testdurchführungsobjektivität besonders dann, wenn die Testgruppe sehr groß ist (z. B. 200-250 Probanden). In manchen Fällen kann das Testresultat dadurch negativ beeinflusst werden – vor allem dann, wenn einige Studierende die Wichtigkeit der Testung in Frage stellen. Die Augenscheinvalidität des Testverfahrens bleibt immer noch sehr wichtig für die Testung, da nur so festgehalten werden kann, dass die Probanden bei den Tests ihre höchste Leistung zeigen. Erscheinen den Studierenden die Gründe für die Testung oder die Testverfahren nicht plausibel oder nicht seriös genug, kann es dazu kommen, dass die Teilnahme an der Testung verweigert wird.

Gewiss sind das nicht alle Gründe, die Einfluss auf die Testergebnisse haben können. Allerdings erschien es an dieser Stelle wichtig, einige der Testprobleme zu beleuchten, die bei der freiwilligen Testung auftreten können.

Die Hypothese, dass die Reihenfolge der Bearbeitung von Tests (C-Tests und des Stolperwörtertests) keine Auswirkungen auf die Testergebnisse hat, wurde nicht bestätigt. Lässt man die Probanden die C-Tests nach dem Stolperwörtertest bearbeiten, werden C-Tests besser gelöst als vor dem Stolperwörtertest.

¹¹⁹ In der VERA-Studie werden Leistungen von Grundschulern in 80-minütigen Tests gemessen und erst im letzten Viertel, d. h. nach 60 Minuten, lassen sich relevante Effekte erfassen.

7.3.4 C-Tests und LT4 (MC-Lesetest und MC-Zuordnungsaufgabe)

Hypothese 7: *Zwischen den durch die C-Tests und durch die MC-Lesetests erfassten Leistungen bestehen Korrelationen.*

Im Folgenden geht es um die Analyse der Kombination aus zwei Testverfahren: C-Tests und MC-Lesetests, die dem Aufgabenpool des TestDaF-Instituts entnommen wurden. Welche Lesetexte für die vorliegende Studie ausgewählt wurden, wurde ausführlich in Kap. 6.4.1 beschrieben (s. Testbeispiele im Anhang A, Teil 2).

Für die Analyse der Daten wird zunächst überprüft, mit welchem Modell sich die Daten besser analysieren lassen. Verwendet werden ein Ratingskalenmodell für die C-Tests (vgl. 7.4.1 und 7.4.2) und ein ein- und zweidimensionales Raschmodell für zwei MC-Lesetests. Im Sinne der Testtheorie bedeutet dies, dass die in der Testbatterie verwendeten Testteile *ein* Testkonstrukt oder *mehrere* Testkonstrukte darstellen, dessen bzw. deren Reliabilität im Weiteren untersucht wird. Im ersten Fall (eindimensionales Raschmodell) wird angenommen, dass der MC-Lesetest und die MC-Leseaufgabe ein gemeinsames Testkonstrukt abbilden (z. B. Überprüfung des globalen Leseverständnisses auf der höheren Prozessebene / Textebene). Im zweiten Fall (zweidimensionales Modell) wird davon ausgegangen, dass die zwei Testverfahren verschiedene Testkonstrukte darstellen und zum Beispiel verschiedene Dimensionen der Lesekompetenz messen.

Verwendung des eindimensionalen Modells:

Bei der Anwendung des eindimensionalen Raschmodells werden der MC-Lesetest und die MC-Zuordnungsaufgabe als ein gemeinsames Konstrukt analysiert. Durch die Analysen hat sich ergeben, dass die Items im MC-Lesetest und in der MC-Zuordnungsaufgabe (insgesamt 20 Items) für die Studierenden eher leicht sind (vgl. auch Kap. 6.4.1). Die Infit-Werte der Items sprechen für eine gute Passung zum Raschmodell. Die berechnete Reliabilität beträgt Cronbachs Alpha $\alpha = 0.788$ und ist für den Untersuchungszweck als ausreichend zu interpretieren. Die Q3-Statistik weist weniger als 10 % auffällige Residualkorrelationen auf.

Verwendung des zweidimensionalen Modells:

Bei der Anwendung des zweidimensionalen Raschmodells werden der MC-Lesetest und die MC-Zuordnungsaufgabe als zwei verschiedene Konstrukte analysiert. Die Infit-Werte der Items sprechen ebenso für eine gute Passung zum Raschmodell. Die Korrelation beider Testteile miteinander (des MC-Lesetests mit der MC-Zuordnungsaufgabe) beträgt $r = 0.643$. Das bedeutet, dass zwischen beiden Testteilen ein mittlerer Zusammenhang besteht. Bei der Durchführung des Modelltests hat sich ergeben, dass das zweidimensionale Modell die Daten signifikant besser beschreibt.

Um den Zusammenhang zwischen den C-Tests und den verwendeten Lesetests zu modellieren, müssen alle drei Tests (C-Tests, MC-Lesetests und MC-Zuordnungsaufgabe) in einem dreidimensionalen Modell abgebildet werden. Als Messmodell für den C-Test wird, wie gehabt, ein Ratingskalenmodell verwendet (vgl. Kap. 7.3.1), als Messmodell für die beiden Lesetests ein Raschmodell. Durch die Analysen ergeben sich folgende Korrelationswerte:

Berechnung der Korrelationen:

Die Korrelationen zwischen den Lesetests (MC-Lesetest und MC-Zuordnungsaufgabe) betragen, wie bereits in einem zweidimensionalen Modell berechnet, $r = 0.643$. Wie erwartet, sind die Korrelationen mit dem zweiten Lesetest (MC-Zuordnungsaufgabe) niedriger, weil der Test in seinem Konstrukt etwas leichter ist als der MC-Lesetest. Die MC-Zuordnungsaufgabe wird in der TestDaF-Prüfung im Bereich Lesen dem Niveau TDN 3 zugeordnet, das von allen drei TDN-Niveaus das niedrigste ist (vgl. dazu Kap. 6.4).

Tabelle 22: Korrelationen zwischen den C-Tests und den MC-Lesetests

verwendete Tests	C-Testbatterie (vier C-Tests)
MC-Lesetest	$r = 0.575$
MC-Zuordnungsaufgabe	$r = 0.397$
beide Lesetests	$r = 0.643$

Wie erwartet, ist der Zusammenhang zwischen den MC-Leseaufgaben unterschiedlich stark. Die Korrelationen zwischen den C-Tests und der MC-

Zuordnungsaufgabe ist aufgrund unterschiedlicher Schwierigkeiten und Aufgabenstellungen geringer als beim MC-Lesetest.

Wie die Daten zeigen, scheint der MC-Lesetest für die Gruppe der Studierenden in Bezug auf sein Niveau angemessener zu sein als die MC-Zuordnungsaufgabe. Wie in Kap. 6.4.1 beschrieben, werden durch den Einsatz von C-Tests und des MC-Lesetests ähnliche Teildimensionen der Lesekompetenz gemessen. So werden sowohl beim C-Test als auch beim MC-Lesetest hierarchiehöhere Prozesse auf der Textebene erfasst (globale Kohärenzbildung, Erkennen von Superstrukturen, Vorwissen, Arbeitsgedächtnis). So kann durch die Ergebnisse darauf geschlossen werden, dass der MC-Lesetest und die C-Tests ähnliche Teilkompetenzen der Lesekompetenz auf ähnliche Weise überprüfen.

Werden beide MC-Lesetests mit den C-Tests korreliert, steigen die Korrelationen an ($r = 0.643$). Dies ist vermutlich dadurch zu erklären, dass die durch die C-Tests erfassten Prozessebenen (Wort-, Satz- und Textebene) in jedem Lesetestteil unterschiedlich hoch repräsentiert sind. So misst zum Beispiel die MC-Zuordnungsaufgabe stärker auf den hierarchieniedrigeren Prozessebenen und der MC-Lesetests mehr auf der Textebene, was sich in der Stärke des Zusammenhangs widerspiegelt.

Die Hypothese, dass zwischen den durch die C-Tests und durch die MC-Lesetests erfassten Leistungen Korrelationen bestehen, wurde bestätigt. Zwischen den C-Tests und den MC-Lesetests bestehen mittlere Korrelationen. Allerdings ist der Zusammenhang beim MC-Lesetest höher als bei der MC-Zuordnungsaufgabe.

Zusammenfassend lässt sich festhalten, dass die C-Tests mit allen in der vorliegenden Studie verwendeten Lesetests korrelieren. Die Höhe des Zusammenhangs lässt sich durch die Aufgabenstellung in den jeweiligen Tests, die Testschwierigkeit und die erfassten Teilkompetenzen sowie durch den Auswertungsmodus erklären. Am schwächsten korrelieren die C-Tests mit dem schulischen Stolperwörtertest ($r = 0.45 - 0.52$) und am höchsten mit dem modifizierten Stolperwörtertest für Erwachsene ($r = 0.52 - 0.82$). Bei der

Kombination von C-Tests und MC-Lesetests wurden niedrige bis mittlere Korrelationen ($r = 0.39 - 0.64$) festgestellt.

Die Kombination der MC-Lesetests bestand (vgl. dazu Kap. 5.4.1) aus zwei verschiedenen Leseverständnistests, einem längeren Lesetext mit einer Multiple-Choice-Aufgabe mit den Antwortoption (*Ja/Nein/Der Text sagt nichts dazu*) und einer MC-Zuordnungsaufgabe, die keinen langen Text darstellt, sondern einzelne kleine Textabschnitte, die bestimmten Aussagen zugeordnet werden müssen. Wie die Korrelationsberechnungen zeigen, besteht zwischen den C-Tests und dem Lesetest mit MC-Fragen zu einem längeren Text ein höherer Zusammenhang als zwischen den C-Tests und der MC-Zuordnungsaufgabe. Dies könnte wie folgt interpretiert werden: Die MC-Zuordnungsaufgabe besitzt für die Gruppe der Studierenden eine geringere Schwierigkeit (vgl. dazu Kap. 5.4) und hat aufgrund dieser Einschränkung in der Gruppe der Studierenden im oberen Leistungsbereich nicht genug Varianz erzeugt, was Auswirkungen auf die Korrelationen hat. Vermutlich fokussiert die MC-Zuordnungsaufgabe stärker andere Teildimensionen der Lesekompetenz, die durch die C-Tests so nicht abgebildet werden können. Durch den Einsatz des Lesetests mit MC-Fragen und die C-Tests werden höhere Korrelationen beobachtet, was darauf schließen lässt, dass beide Testinstrumente ähnliche Teilkompetenzen erfassen (z. B. Kohärenzherstellung auf der Satz- und Textebene).

Zwischen allen mit den C-Tests zusammen verwendeten Lesetests wurden schwache bis hohe Korrelationen beobachtet, was als ein Indiz dafür interpretiert werden kann, dass sich durch die C-Tests verschiedene Teildimensionen der Lesekompetenz abbilden lassen (vgl. Kap. 2 zu Lesekompetenz und Lesemodellen). Auch wenn C-Tests die Lesefähigkeiten nicht auf ähnliche Weise abbilden wie dies z. B. durch den Einsatz von Lesetests mit MC-Fragen möglich ist, so hat die vorliegende Studie doch gezeigt, dass die C-Tests Lesefähigkeiten erfassen. Die berechneten Korrelationen zeigen, dass die C-Tests auf den hierarchieniedrigeren und -höheren Prozessebenen differenziert messen können. Wird ein anderer Test (Lesetest) zur Überprüfung der Prozesse eingesetzt (z. B. auf der Wort- und Satzebene), dann korrelieren beide

Testinstrumente stärker auf der Ebene dieser Teildimensionen miteinander. Sind die Korrelationen schwach, kann man davon ausgehen, dass entweder a) beide Tests verschiedene Prozessebenen erfassen, die miteinander wenig zu tun haben, b) die Fähigkeiten auf unterschiedliche Weise erfasst werden oder c) die verwendeten Testinstrumente unterschiedlich schwierig sind (z. B. der eine Test ist leichter als der andere).

Die festgestellten Korrelationen belegen, dass die eingesetzten Tests (C-Tests und verschiedene Lesetests) nur bis zu einem gewissen Grad ähnliche Teilkompetenzen erfassen. Nicht bei allen in Kombination mit den C-Tests verwendeten Lesetests sind diese Teilkompetenzen ähnlich. Die Höhe des Zusammenhangs und die Beschreibung der Prozesse, die bei einzelnen Lesetests aktiviert werden, zeigen, inwieweit das ausgewählte Testformat zur Erfassung der Teildimensionen der Lesekompetenz eingesetzt werden kann (zur Beschreibung einzelner Testkonstrukte s. Kap. 5.4.1 und 5.4.2). Die festgestellten hohen Korrelationen zwischen den C-Tests und dem Stolperwörtertest (vgl. dazu Kap. 7.3.2) tragen auch dazu bei, dass an dieser Stelle gesagt werden kann, dass der C-Test in Bezug auf die Erfassung von Lesefähigkeiten als ein valides Instrument im Sinne der Vorhersagevalidität (vgl. dazu Kap. 8.2) angesehen werden kann. Die berechneten Korrelationen belegen, dass der C-Test als Prädiktor zur Voreinschätzung der Leseleistungen der Studierenden in der Praxis eingesetzt werden kann.

8. Hypothesen: Ergebnisse und Kommentare

In diesem Kapitel werden die Gesamtergebnisse mit besonderem Augenmerk auf die C-Tests dargestellt. Für die Analyse und Überprüfung formulierter Hypothesen werden zusätzliche Daten verwendet, die im Rahmen der Testungen mithilfe der Fragebögen erfasst wurden und in Kap. 8.1 beschrieben werden (s. Fragebögen im Anhang A, Teil 1). Die Hypothesen beziehen sich auf persönliche Daten der Studierenden wie z. B. Geschlecht, Alter oder Schulabschluss. In Kap. 8.2 werden Hypothesen dargestellt, die in Bezug auf die Testergebnisse differenzierte Aussagen erlauben können, falls die Hypothesen gelten. So werden z. B. selbst eingeschätzte Kompetenzen (globale Einschätzung der Sprachkompetenz, Lesekompetenz und schriftliche Kompetenz) mit den durch die C-Tests oder Lesetests eingeschätzte Leistungen korreliert.

8.1.1 Hypothesen zu den Hintergrundvariablen

8.1.1 C-Testleistung und Geschlecht

Hypothese Nr. 8: *Sprachliche Kompetenzen männlicher und weiblicher Studierender unterscheiden sich.*

Angenommen wird, dass weibliche Studierende höhere C-Testergebnisse erzielen als männliche Studierende.

Über unterschiedliche Schulleistungen männlicher und weiblicher Probanden in den sprachlichen Fächern berichten mehrere Leistungsstudien (z. B. PISA 2000, OECD 2004, DESI 2008). Die Ergebnisse der PISA-Studie (2000) zeigten, dass Mädchen im Lesen signifikant höhere Testwerte erreichen als Jungen. Besonders ausgeprägt ist der Leistungsvorsprung der Mädchen bei kontinuierlichen Texten, während bei nichtkontinuierlichen Texten sehr viel geringere Geschlechtsunterschiede zu verzeichnen sind (vgl. Kap. 1.6.1). Auch im Hinblick auf die Anforderungen an den Umgang mit Texten zeigen sich geschlechtsspezifische Unterschiede: Im Vergleich zu Mädchen bereitet es

8. Hypothesen: Ergebnisse und Kommentare

Jungen deutlich größere Schwierigkeiten, Texte und ihre Merkmale kritisch zu reflektieren und zu bewerten (vgl. PISA-Konsortium, 2000, S. 257).

Von ähnlichen Befunden berichtet auch die DESI-Studie (2008) bei jüngeren Schülerinnen und Schülern. Der Leistungsvorsprung am Ende des sprachlichen Entwicklungsprozesses der Mädchen manifestiert sich in der Grundschulzeit besonders im Bereich der Lesekompetenz sowie in der Rechtschreibung. Mädchen zeigen eine schnellere Entwicklung von Rechtschreibstrategien und eine generell höhere Lernmotivation im sprachlichen Kompetenzerwerb (vgl. Hartig & Jude, 2008, S. 202). Die DESI-Ergebnisse aus dem Fach Deutsch verzeichnen eine Leistungsdifferenz von 40 Punkten, was einem mittleren Effekt ($d = 0.41$) entspricht, von dem auch in der PISA-Studie (2003) für die Lesekompetenz ($d = 0.39$) berichtet wurde. In der genannten Studie wurde in allen erhobenen Kompetenzbereichen für Deutsch ein statistisch bedeutsamer Vorsprung der Mädchen beobachtet. Die deutlichsten Geschlechtsdifferenzen sind in den produktiven Teilkompetenzen (Textproduktion und Rechtschreibung) sowie in der Argumentation zu beobachten (vgl. DESI-Konsortium, 2008, S. 203). Bei höheren Altersstufen (z. B. beim Übergang zur Universität) wurden die sprachlichen Leistungsunterschiede bis jetzt nicht untersucht. Ausgehend von den Ergebnissen der genannten Studien wird im Rahmen dieser Untersuchung angenommen, dass Geschlechtsdifferenzen auch im universitären Bereich bestehen und weibliche Studierende bessere C-Testleistungen erzielen als männliche Studierende. Für die erfolgreiche Rekonstruktion von C-Tests wird die Lesekompetenz vorausgesetzt, die sich in der Anwendung verschiedener Texterschließungsstrategien niederschlägt. Dass die C-Testergebnisse und die Ergebnisse verschiedener Leseaufgaben miteinander korrelieren, wurde im fremdsprachlichen Bereich untersucht. Im Rahmen dieser Studie wurden Korrelationen zwischen den muttersprachlichen C-Tests und den Lesetests in Deutsch überprüft (vgl. Kap. 6, Kap. 7).

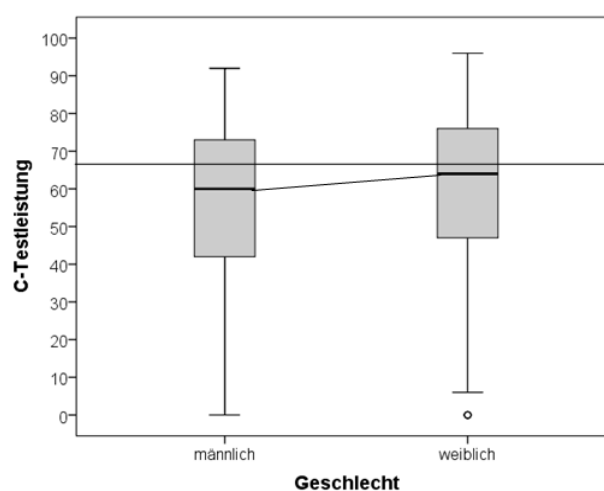
Die Ergebnisse der Untersuchung werden im Weiteren mit Hilfe eines *Box-Plots* dargestellt (s. Abb. 14). Ein Boxplot besteht aus einem Rechteck (Box) und zwei Linien, die über das Rechteck hinausweisen. Eine schwarze, fettmarkierte Linie in der Mitte der Box ist der *Median*, der als jener Wert interpretiert wird, der die

8. Hypothesen: Ergebnisse und Kommentare

Stichprobe in zwei gleiche Hälften zu je 50 % teilt (vgl. Bühner, 2004, S. 74). Die über die Boxen nach oben und unten hinausweisenden Linien werden als *Whiskers* bezeichnet und werden durch einen Strich abgeschlossen. Sie kennzeichnen die Bereiche, in denen die übrigen 50 % der Untersuchungsfälle liegen. Der im Vergleich zum arithmetischen Mittel (Mittelwert) gegen Ausreißer robustere Median unterteilt die Box noch einmal in zwei jeweils 25 % der Ergebnisse der Gesamtgruppe repräsentierende Teilbereiche oder *Quartile*.

Vergleicht man die Verteilungen der C-Testleistungen innerhalb der Gruppen miteinander, so ist auffällig, dass zum einen die Varianzen innerhalb der Gruppen unterschiedlich groß sind und zum anderen die Medianwerte voneinander abweichen. Weibliche Studierende, die in der vorliegenden Untersuchung überrepräsentiert sind (N = 665), was für die Spezifik unserer Untersuchungsgruppe zu erwarten ist, da mehr Frauen den Lehrerberuf wählen, erzielen im Durchschnitt *signifikant bessere C-Testleistungen* als männliche Studierende (N = 219) (vgl. Berechnungen im Anhang A, Teil 3.3.1). Die durchschnittliche C-Testleistung bei weiblichen Studierenden beträgt 61 Punkte und der Medianwert 64 Punkte. Bei männlichen Studierenden beträgt die durchschnittliche Testleistung nur 57 Punkte und der Medianwert 60 Punkte.

Abb. 14: C-Testergebnisse männlicher und weiblicher Studierender (Gesamtgruppe)



Da die vorliegende Untersuchungsgruppe, die aus Lehramtsstudierenden im Grundstudium besteht, der Datengrundlage aus dem *Sprachkompetenzprojekt*

8. Hypothesen: Ergebnisse und Kommentare

Förderung und Diagnose an der UDE ähnelt und somit die gleiche Population der Studierenden darstellt, wurden für die vorliegende Untersuchung die im Projekt festgelegten Leistungsstufen und der *Referenznormwert* übernommen (s. Tabelle 35). Der Referenznormwert wird als ein standardisierter Testwert verstanden, der an der Gruppe der monolingualen Studierenden mit der Muttersprache Deutsch berechnet wird. Er entspricht der bereinigten durchschnittlichen C-Testleistung (zum Begriff und seine statistisch-normative Festlegung vgl. Baur, Grotjahn & Spettmann, 2006).

Verbindet man in der Grafik die Medianwerte beider Gruppen miteinander zu einer Linie, wird sichtbar, dass die von beiden Gruppen erreichten Werte unter dem für die Gruppe der Lehramtsstudierenden ermittelten Referenznormwert liegen, der 67 Punkte beträgt¹²⁰. Das bedeutet, dass die Mehrheit der Studierenden in beiden Gruppen im Vergleich zum normierten durchschnittlichen Testwert schwächere Testleistungen produziert haben. Die C-Testleistung, die der Leistungsstufe IV oder V zugeordnet wird, wird als eine schwache bzw. sehr schwache Leistung interpretiert. Studierende, die auf diesen Stufen positioniert sind, haben nach den C-Testergebnissen einen großen oder sogar erheblichen sprachlichen Förderbedarf (zur Feststellung von Leistungsstufen und deren Interpretation vgl. Scholten-Akoun & Baur, 2011; Scholten-Akoun, Kuhnen & Mashkovskaya, 2012; Scholten-Akoun, Tischmeyer & Mashkovskaya, 2013).

Zur Überprüfung der Verteilung der männlichen und weiblichen Studierenden innerhalb verschiedener Leistungsstufen werden die Testergebnisse aufgrund des erreichten Wertes einer der fünf Leistungsstufen zugeordnet. Wie die Tabelle 35 zeigt, sind männliche Studierende in den oberen Leistungsstufen I und II nur mit 10,9 % vertreten und im Gegensatz zu den weiblichen Studierenden, deren Anteil 16,4 % der Gesamtgruppe der weiblichen Studierenden in diesen oberen Bereichen ausmacht, unterrepräsentiert. Auf den unteren Leistungsstufen dagegen sind männliche Studierende mit 25,6 % für die Leistungsstufe IV und mit 14,6 % für die Leistungsstufe V überrepräsentiert. Ein

¹²⁰ Die Festlegung der Leistungsbereiche bzw. Leistungsstufen wurde aus dem *Sprachkompetenzprojekt Diagnose & Förderung* an der Universität Duisburg-Essen übernommen (vgl. Publikation von Bremerich-Vos et al., 2013).

8. Hypothesen: Ergebnisse und Kommentare

ähnliches Muster findet sich in der OECD-Studie bei verschiedenen Teilnehmerstaaten wieder (vgl. OECD, 2011).

Tabelle 23: Anteil von weiblichen und männlichen Probanden auf den Leistungsstufen in den C-Tests (Gesamtgruppe), N = 887

	Leistungs- stufe I	Leistungs- stufe II	Leistungs- stufe III	Leistungs- stufe IV	Leistungs- stufe V	Gesamt
männlich	4 1,8 %	20 9,1 %	107 48,9 %	56 25,6 %	32 14,6 %	219 24,7 %
weiblich	32 4,8 %	77 11,6 %	351 52,7 %	130 19,5 %	76 11,4 %	666 75,1 %
Gesamt	36 4,1 %	97 10,9 %	458 51,6 %	186 21,0 %	110 12,4 %	887 100 %

Legende: Leistungsstufe I: > als 89 Punkte; Leistungsstufe II: 82-89 Punkte; Leistungsstufe III: 53-81 Punkte; Leistungsstufe IV: 36-52 Punkte; Leistungsstufe V: < als 35 Punkte

Partialisiert man die Gruppe der Studierenden mit einer anderen Muttersprache als Deutsch heraus, so ergeben sich in der Gruppe der deutschen Muttersprachler ebenfalls starke Unterschiede in der C-Testleistung zwischen männlichen und weiblichen Probanden. Auch hier erreichen weibliche Studierende (N = 397) im Durchschnitt signifikant höhere C-Testleistungen als ihre männlichen Kommilitonen (N = 149). Der Mittelwert in der Gruppe der weiblichen Studierenden mit der Muttersprache Deutsch beträgt 70 Punkte und in der Gruppe der männlichen Studierenden nur 63 Punkte, $p = .000 < .01$ (Sig. 2-seitig). Ob diese Unterschiede sich auch in der Gruppe der Studierenden mit Migrationshintergrund beobachten lassen, wird in Hypothese 2 überprüft.

Die Hypothese, dass weibliche Studierende höhere C-Testergebnisse erzielen als männliche Studierende, konnte bestätigt werden.

8.1.2 C-Testleistung und Migrationshintergrund

Hypothese 9: *Studierende mit Migrationshintergrund erzielen im C-Test schlechtere Ergebnisse als Studierende ohne Migrationshintergrund.*

Viele Leistungsstudien und amtliche Statistiken belegen, dass ein substanzieller Teil der schulpflichtigen Schüler und Schülerinnen aus Familien stammt, in denen

zumindest ein Elternteil nicht in Deutschland geboren wurde und eine andere Muttersprache als Deutsch spricht. Die Schule ist ein guter Spiegel dieses Trends, da „multi-ethnisch zusammengesetzte Klassen“ in vielen Schulen heute fast die Regel sind (PISA-Konsortium, 2000, S. 340). Amtliche Statistiken können nicht immer Auskunft darüber geben, ob ein Migrationshintergrund besteht (z. B. im Falle einer Einbürgerung bei gemischten Ehen, vgl. PISA-Konsortium, 2000, S. 340) und welche Sprachen in der Familie gesprochen werden. Weder in der Schule noch in anderen Bildungseinrichtungen werden diese Daten systematisch erhoben (vgl. Chlosta & Ostermann, 2005; Scholten-Akoun & Baur, 2012; Scholten-Akoun, Kuhnen & Mashkovskaya, 2012).

Die im Zentrum des Interesses stehende Variable Migrationshintergrund wurde in der vorliegenden Studie durch Fragen zur Muttersprache beider Eltern erfasst. Gaben die Probanden an, dass ein Elternteil eine andere Muttersprache als Deutsch spricht, wurde dieser Studierende in die Gruppe mit Migrationshintergrund aufgenommen¹²¹. Insgesamt wurden in den Projekterhebungen 38 verschiedene Sprachen genannt, die in den Familien als Muttersprache gesprochen werden¹²².

In der Gruppe der getesteten Studentenschaft gab etwa ein Drittel (37,5 %) der Studierenden (N = 333) an, dass mindestens ein Elternteil eine andere Muttersprache spricht als Deutsch. Im Weiteren werden diese Studierenden als Studierende mit Migrationshintergrund (MH) bezeichnet. Die größten Sprachgruppen bilden in der vorliegenden Untersuchung türkisch-, russisch- und polnischsprachige Studierende, auf die im Folgenden genauer eingegangen wird. Die gleichen Gruppen überwiegen auch in der DESI-Studie (DESI-Konsortium, 2008, S. 211) und in der Untersuchungsgruppe des *Sprachkompetenzprojektes Diagnose und Förderung* (UDE) (vgl. Bremerich-Vos et al., 2013).

¹²¹ Bei der gängigen statistischen Erfassung des Migrationshintergrundes wird u. a. auch noch der Geburtsort der Testteilnehmenden und/oder der Eltern erfasst (vgl. Stanat, Rauch & Segeritz 2010, S. 204), was im Rahmen dieser Untersuchung nicht geschah.

¹²² Albanisch, Arabisch, Aramäisch, Armenisch, Berbisch, Bosnisch, Chinesisch, Dari, Deutsch, Englisch, Farsi, Flämisch, Georgisch, Griechisch, Hindi, Indonesisch, Italienisch, Japanisch, Jugoslawisch, Kirgisisch, Koreanisch, Kroatisch, Kurdisch, Marokkanisch, Niederländisch, Polnisch, Russisch, Serbisch, Serbo-Kroatisch, Slowakisch, Spanisch, Tamilisch, Thailändisch, Tschechisch, Türkisch, Turkmenisch, Ukrainisch, Ungarisch.

Tabelle 24: Die größten Sprachgruppen in der Gesamtgruppe (N = 887)

Türkisch 157 (17,7 %)	Russisch 39 (4,4 %)	Polnisch 27 (3 %)
Türkisch & Türkisch 5 (0,6 %)	Russisch & Russisch 37 (4,2 %)	Polnisch & Polnisch 26 (2,9 %)
Türkisch & Deutsch 5 (0,6 %)	Russisch & Deutsch 2 (0,2 %)	Polnisch & Deutsch 1 (0,1 %)

*Die Angaben in der Tabelle betreffen die Muttersprache beider Elternteile

Wie aus Tabelle 24 ersichtlich, bilden Studierende mit einem türkischen Migrationshintergrund die größte Gruppe, gefolgt von Studierenden mit einem russischen und danach mit einem polnischen Migrationshintergrund. Die Fragen, ob beide Sprachen zur gleichen Zeit erworben wurden und ob diese in der Familie der Testteilnehmenden als Kommunikationssprachen genutzt werden, in welchem Alter Studierende nach Deutschland eingewandert sind, falls sie in einem anderen Land geboren wurden oder welchen sozial-ökonomischen Status ihre Familie hat, wurden in der vorliegenden Studie nicht gestellt¹²³.

Die Angaben der Studierenden bezüglich der Muttersprache liefern in Bezug auf die sprachlichen Leistungen insofern wichtige Hinweise, als im aktuellen Bildungsdiskurs Studierenden mit Migrationshintergrund ein privilegierter Zugang zum Lehramtsstudium gewährt werden soll (vgl. Baur & Scholten, 2010; Scholten-Akoun & Baur, 2011)¹²⁴. Nordrhein-westfälische Schulen sollen mehr LehrerInnen mit Migrationshintergrund aufnehmen, um unter anderem die sprachliche Förderung derjenigen Schüler zu stärken, die ebenfalls einen Migrationshintergrund haben. Dies aber kann aufgrund der in vielen Studien erfassten unterdurchschnittlichen sprachlichen Leistungen von Schulabsolventen mit Migrationshintergrund zu Problemen führen:

„Rekrutierungskampagnen wie 'Mehr Migranten werden Lehrer' oder politische Forderungen z. B. der Integrationsbeauftragten der Bundesregierung, Maria Böhmer, die einen Anteil von MigrantInnen im öffentlichen Dienst fordert, der in etwa dem Bevölkerungsanteil von 20 % entspricht ('Besonders dringend benötigen wir mehr Lehrerinnen und

¹²³ Diese Daten wurden zum Beispiel bei Grundschulern im Rahmen des Essener SPREEG-Projekts erhoben: http://www.uni-due.de/unikate/ressourcen/grafiken/PDF's/EU_24/24_baur.pdf, zuletzt gesehen am 13.06.2013.

¹²⁴ http://www.fadaf.de/de/daf_angebote/daz/daz_in_der_lehrerausbildung_juli_2010.pdf

8. Hypothesen: Ergebnisse und Kommentare

Lehrer sowie Erzieherinnen mit Migrationshintergrund') scheinen ohne begleitende Investitionen in die Verbesserung der basalen sprachlichen Kompetenzen der Geförderten als fragwürdige Instrumente zur Verbesserung der Integration und der Förderung gerade von Kindern mit Migrationshintergrund an den Schulen.“ (Scholten-Akoun & Baur, 2011, S. 326)

Die Autoren fragen zu Recht, warum zukünftige LehrerInnen mit eigenem Migrationshintergrund gleichsam „natürlich die besseren sprachlichen Förderer der Kinder mit Migrationshintergrund“ sein sollen. Scholten-Akoun und Baur betonen für Lehramtsstudierende die Notwendigkeit der Beherrschung des Deutschen auf einem für den Lehrerberuf geeigneten Niveau, da „ohne eine eigene solide sprachliche Basis“ weder Lehrerinnen mit noch Lehrerinnen ohne Migrationshintergrund ihre sprachfördernden Aufgaben angemessen wahrnehmen können (Scholten-Akoun & Baur, 2011, S. 326).

Da heftige Diskussionen über den Erhalt und die Förderung von Zwei- und Mehrsprachigkeit von Migranten und die damit verbundene Entwicklung der sprachlichen Fähigkeiten der Schüler mit Migrationshintergrund geführt werden (vgl. Gogolin & Neumann, 2009, Streitfall Zweisprachigkeit - *The Bilingualism Controversy*), stellt sich die Frage, ob und wie stark die sprachlichen Leistungen der beiden Gruppen (Studierende mit und ohne Migrationshintergrund) sich voneinander unterscheiden. Im Folgenden wird dieser Aspekt in der zu untersuchenden Gruppe der Lehramtsstudierenden dargestellt.

In der vorliegenden Untersuchungsgruppe erzielen Studierende ohne Migrationshintergrund im Durchschnitt *signifikant höhere Testergebnisse* (66 Punkte) als Studierende mit Migrationshintergrund (50 Punkte) (vgl. Abb. 15). In Abbildung 15 sind die Differenzen zwischen den Gruppen gut zu erkennen; zum Beispiel ist die Streuung der Leistungen innerhalb der Gruppen unterschiedlich breit, vor allem im unteren Leistungsbereich in der Gruppe der Studierenden mit Migrationshintergrund. Dies bedeutet, dass diese Studierenden u. a. sehr wenige Punkte in den C-Tests erreicht haben (vgl. auch Berechnungen im Anhang A, Teil 3.3.2).

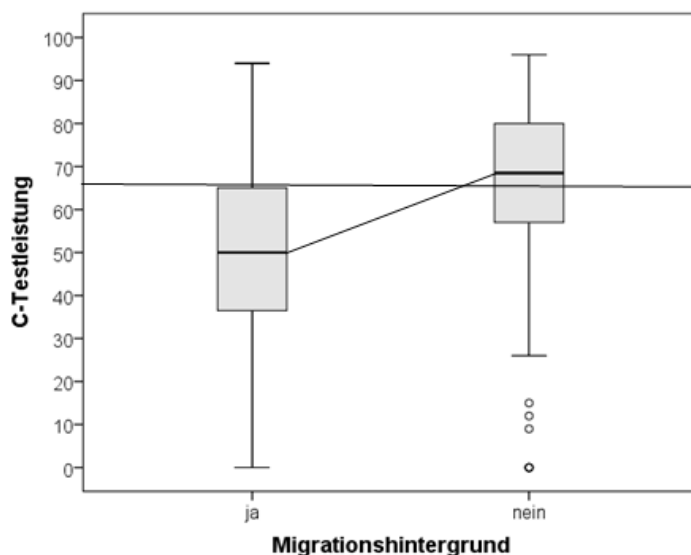
8. Hypothesen: Ergebnisse und Kommentare

Zu ähnlichen Ergebnissen kamen Scholten-Akoun & Baur (2011), die anhand der Daten aus dem Sprachkompetenzprojekt der Universität Duisburg-Essen belegten,

„... dass Studierende mit Migrationshintergrund *im Durchschnitt* ihr Lehramtsstudium mit schlechteren sprachlichen Voraussetzungen beginnen – darauf sei noch einmal eigens hingewiesen –, bedeutet nicht, dass etwa alle individuellen Leistungen unter denen der Studierenden ohne Migrationshintergrund lägen oder dass etwa die über die C-Test-Ergebnisse bestimmte Risikogruppe (LB 4 und 5) mit nur wenigen Ausnahmen aus Studierenden mit Migrationshintergrund bestünde. Das Leistungsspektrum beider Gruppen lässt sich innerhalb vergleichbarer Grenzen beschreiben und zur Risikogruppe zählen Studierende mit und ohne Migrationshintergrund.“ (Scholten-Akoun & Baur, 2011, S. 325)

Verbindet man die Medianwerte beider Gruppen grafisch zu einer Linie, so zeigt sich eine sehr große Differenz. Der Medianwert der Studierenden mit Migrationshintergrund liegt deutlich unter dem Wert der Studierenden ohne Migrationshintergrund und deutlich unter dem ermittelten Referenznormwert.

Abb. 15: C-Testleistungen der Studierenden mit und ohne Migrationshintergrund



In der Grafik ist gut zu erkennen, dass im Vergleich zu der Gruppe der Studierenden ohne Migrationshintergrund innerhalb der Gruppe der Studierenden mit Migrationshintergrund eine sehr viel größere Varianz vorliegt. Betrachtet man die Ergebnisse der Studierenden mit Migrationshintergrund genauer, so ergibt sich Folgendes: Rund 25 % der Studierenden erzielten Werte,

8. Hypothesen: Ergebnisse und Kommentare

die über dem Referenznormwert liegen, und ca. 75 % der C-Testleistungen liegen unter dem Referenznormwert. Vergleicht man diese Werte mit den Leistungen der Studierenden ohne Migrationshintergrund, so stellt sich heraus, dass in dieser Gruppe mehr als die Hälfte der Testleistungen über dem Referenznormwert liegt.

Teilt man Studierende mit und ohne Migrationshintergrund in verschiedene Leistungsgruppen bzgl. des Referenznormwertes auf, so kommt man zu folgendem Ergebnis: Studierende mit Migrationshintergrund sind in den unteren Stufen IV und V überrepräsentiert, auf der Leistungsstufe V sogar mit 23,7 % gegenüber 5,1 % in der Gruppe der deutschen Muttersprachler. Diese erhobenen Daten entsprechen den Befunden vieler Studien (z. B. PISA, OECD, DESI), die für die untersuchte Gruppe der Testteilnehmenden mit Migrationshintergrund ähnliche Ergebnisse für die Deutschkenntnisse dokumentieren.

Tabelle 25: C-Testleistungen der Studierenden mit und ohne Migrationshintergrund, aufgeteilt nach Leistungsstufen

Studierende	Leistungsstufe I	Leistungsstufe II	Leistungsstufe III	Leistungsstufe IV	Leistungsstufe V	Gesamt
mit MH	3 0,9 %	13 3,9 %	133 39,9 %	105 31,5 %	79 23,7 %	333 37,5 %
ohne MH	32 5,9 %	84 15,4 %	324 59,3 %	78 14,3 %	28 5,1 %	546 61,5 %
Gesamt	36 4,1 %	97 10,9 %	458 51,6 %	186 21,0 %	110 12,4 %	887 100 %

Legende: Leistungsstufe I: > als 89 Punkte; Leistungsstufe II: 82-89 Punkte; Leistungsstufe III: 53-81 Punkte; Leistungsstufe IV: 36-52 Punkte; Leistungsstufe V: < als 35 Punkte

Wenn man die Gruppe der Studierenden mit Migrationshintergrund nach verschiedenen Sprachgruppen aufteilt, kann man feststellen, dass in allen drei Gruppen nur wenige Studierende die höheren Leistungsstufen I und II erreichen. Die Leistungen der meisten Probanden liegen unter dem Referenznormwert (unter der Leistungsstufe III). Allerdings muss hier beachtet werden, dass die Gruppen unterschiedlich groß sind. Die größte Gruppe bilden türkischsprachige Studierende, was für die Region, in der die Studie durchgeführt wird (Ruhrgebiet)

8. Hypothesen: Ergebnisse und Kommentare

zu erwarten ist. Neben der Tatsache, dass das Ruhrgebiet seit den fünfziger Jahren eine der großen Einwanderungsregionen für türkischsprachige Familien ist, muss berücksichtigt werden, dass es an der Universität Essen (seit 2003 Duisburg-Essen) ein einzigartiges Studienangebot für das Fach Türkisch gibt. Seit Mitte der neunziger Jahre kann hier das Fach Türkisch als Lehramtsfach studiert werden, wodurch diese Universität für viele türkischsprachige Studierende besonders attraktiv geworden ist¹²⁵.

In der Gesamtgruppe erreichen nur sehr wenige Probanden die höheren Leistungsstufen I und II. Die Leistungen von mehr als der Hälfte aller Studierenden in der türkisch- und russischsprachigen Gruppe ist den unteren Leistungsstufen (IV und V) zuzuordnen; in der polnischsprachigen Gruppe sind es 44,4%.

Tabelle 26: C-Testleistungen, aufgeteilt nach Leistungsstufen und Sprachgruppen

Studierende	Leistungsstufe I	Leistungsstufe II	Leistungsstufe III	Leistungsstufe IV	Leistungsstufe V	Gesamt
Türkisch	-	4 2,5 %	67 42,7 %	53 33,8 %	33 21,0 %	157 71,0 %
Polnisch	1 3,7 %	2 7,4 %	12 44,4 %	10 37 %	2 7,4 %	27 12,2 %
Russisch	-	1 2,7 %	15 40,5 %	16 43,2 %	7 18,9 %	37 16,7 %
Gesamt	1 0,5 %	7 3,2 %	94 42,5 %	79 35,7 %	42 19,0 %	221 100 %

Legende: Leistungsstufe I: > als 89 Punkte; Leistungsstufe II: 82-89 Punkte; Leistungsstufe III: 53-81 Punkte; Leistungsstufe IV: 36-52 Punkte; Leistungsstufe V: < als 35 Punkte

Zusammenfassend lässt sich festhalten, dass die Gruppe der Studierenden mit Migrationshintergrund signifikant schwächere Ergebnisse in der C-Testleistung erzielt. Dies lässt vermuten, dass diese Unterschiede, die mit den C-Tests global erfasst wurden, auch in einzelnen sprachlichen Bereichen existieren (z. B. in schriftsprachlichen Teilkompetenzen). In welchen sprachlichen Bereichen Schwächen bei der Migrantengruppe (und natürlich auch bei monolingualen

¹²⁵ Startseite des Instituts für Turkistik: <http://www.uni-due.de/turkistik/>, zuletzt gesehen am 05.06.2013.

8. Hypothesen: Ergebnisse und Kommentare

Studierenden in der Gesamtgruppe) stärker vorkommen, soll mit anderen diagnostischen Instrumenten erhoben werden, die aufgrund einer kriteriumsorientierten Bewertung der Stärken und Schwächen Studierender mit und ohne Migrationshintergrund ausführlichere Informationen liefern können. Im Rahmen dieser Studie wird der Bereich Lesen mit entsprechenden Lesetests erfasst. Die Ergebnisse dazu werden in Kap. 7 dargestellt.

Wie die Testergebnisse bestätigen, benötigt vor allem die Gruppe der Studierenden mit Migrationshintergrund eine genauere Analyse ihrer sprachlichen Leistungen, da ihre Testergebnisse signifikant häufiger den unteren und untersten Leistungsstufen zuzuordnen sind.

Zum einen wird ein spezifischeres diagnostisches Instrument in Form von gezielten Aufgaben benötigt, die genaue Informationen über die zu fördernden Bereiche liefern können, und zum anderen sollten geeignete Fördermaßnahmen angeboten werden, die an die Bedürfnisse dieser Gruppe angepasst sind und an eine solche Testung anknüpfen können.

Die Schulleistungsstudien der letzten Jahre (OECD, 2003; DESI-Konsortium, 2008) haben gezeigt, dass die schulischen und sprachlichen Leistungen von Kindern und Jugendlichen mit Migrationshintergrund im Durchschnitt schwächer sind als die Leistungen monolingualer Mitschüler. Im Bereich der sprachlichen Leistungen von Lehramtsstudierenden lässt dies sich jetzt erstmalig auch im universitären Kontext bestätigen. Welche Faktoren die sprachlichen Leistungen der Studierenden mit Migrationshintergrund beeinflussen, wurde im Rahmen dieser Studie nicht untersucht. Die Interpretation der Testergebnisse einzelner Probanden und spezifischer Teilgruppen ist ohne Einbeziehung von Einflussfaktoren (z. B. der Lernumgebung, in der die Studierenden leben) oder des sozio-ökonomischen Status' der Familie nicht möglich. Ungeachtet verschiedener Faktoren und Variablen ist es eine Tatsache, dass für das Erreichen guter sprachlicher und fachsprachlicher Leistungen sowohl in der Schule als auch in höheren Bildungsinstitutionen wie der Universität für Kinder, Jugendliche und Studierende mit Migrationshintergrund Fördermaßnahmen bereitgestellt

werden müssen, um die Fähigkeiten dieser Gruppe als gesellschaftliche Ressource zu entwickeln.

Die Hypothese, dass Studierende mit Migrationshintergrund im C-Test schlechtere Ergebnisse als Studierende ohne Migrationshintergrund erzielen, konnte bestätigt werden.

8.1.3 C-Testleistung und Alter

***Hypothese 10:** Es bestehen Unterschiede zwischen den C-Testleistung jüngerer und älterer Probanden innerhalb der Gruppe der getesteten Studierenden.*

Angenommen wird, dass sprachliche Leistungen, die durch den C-Test gemessen werden können, bei älteren Studierenden besser sind als bei jüngeren Studierenden.

Die Entwicklung der allgemeinen Sprachkompetenz und ihrer einzelnen Teildimensionen ist mit Eintritt ins Erwachsenenalter nicht abgeschlossen. Teildimensionen der Sprachkompetenz wie z. B. mit dem Lesen und Schreiben verbundene fachspezifische rezeptive und produktive Fähigkeiten werden im Laufe des Studiums weiterentwickelt.

Da an der Universität die Gruppe der Studierenden im Grundstudium in Bezug auf das Alter nicht homogen ist und einige Studierende sogar bereits ein abgeschlossenes Studium bzw. eine abgeschlossene Ausbildung haben, wird hier der Frage nachgegangen, ob die Sprachkompetenz mit zunehmendem Alter steigt und sich in höheren C-Testleistungen widerspiegelt. Angenommen wird, dass entsprechende sprachlichen Unterschiede (falls vorhanden) auch mithilfe von C-Tests messbar sind.

Der Analyse der sprachlichen Leistung von Jugendlichen und jungen Erwachsenen widmete sich Wockenfuß in einer Studie (2009)¹²⁶. Sie überprüfte die globale Sprachfähigkeit von Jugendlichen und jungen Erwachsenen mit den C-Tests und stellte fest, dass das Wachstumsmodell für den Entwicklungsaufbau der globalen Sprachfähigkeit nur bis zum 25. Lebensjahr zu beobachten ist. Ab

¹²⁶ In der Untersuchung von Wockenfuß wurden die Deutschnoten von 988 Probanden erhoben, von denen 814 Probanden im Alter von 9 bis 19 Jahren und 174 Probanden im Alter von 20 bis 57 Jahren waren.

8. Hypothesen: Ergebnisse und Kommentare

dem 25. Lebensjahr und danach wurde dagegen eine Abnahme der C-Testleistungen festgestellt (vgl. Wockenfuß, 2009, S. 190f.).

In der vorliegenden Studie wurden Alterskategorien gebildet und die C-Testleistung der Studierenden, die verschiedenen Altersgruppen angehören, analysiert. Das Alter des jüngsten Probanden in der Studie beträgt 19 Jahre und das des ältesten 64 Jahre.

Am häufigsten kommen 20-jährige (165 Probanden; 18,6 %), 21-jährige (176 Probanden; 19,8 %) und 22-jährige Studierende (129 Probanden; 14,5 %) vor. Das durchschnittliche Alter der Studierenden beträgt 23,6 Jahre (SD = 4,8 Jahre). Um die C-Testleistungen auf verschiedenen Altersstufen zu beobachten, wurden fünf Altersgruppen gebildet.¹²⁷

Tabelle 27: C-Testleistungen in verschiedenen Alterskategorien

Alterskategorien	Häufigkeiten	C-Testleistung (Mittelwert, Median)	min. und max. C-Testleistung
I. 19 bis 25 Jahre	698 (78,7 %)	59,9 (63)	min. 0/max. 96
II. 26 bis 30 Jahre	103 (11,6 %)	61,2 (66)	min. 0/max. 94
III. 31 bis 35 Jahre	46 (5,2 %)	60,7 (65)	min. 0/max. 92
IV. 36 bis 40 Jahre	15 (1,7 %)	70,1 (72)	min. 35/max. 91
V. älter als 40 Jahre	11 (1,2 %)	57,9 (64)	min. 9/max. 89

Aus Tabelle 27 ist ersichtlich, dass ein kleiner Anstieg der durchschnittlichen Leistung zwischen der ersten und zweiten Kategorie vorliegt. In der dritten Alterskategorie ist ein geringer Leistungsabfall zu beobachten. Aussagekräftigere Informationen bezüglich der erzielten C-Testleistung in den Gruppen liefern hier die Medianwerte (vgl. die dritte Spalte in der Tabelle). Der höchste Medianwert ist in der Gruppe der 36- bis 40-jährigen Studierenden festzustellen. Er beträgt 72 Punkte und liegt damit deutlich über dem ermittelten Referenznormwert (67 Punkte). Allerdings sind dieser Kategorie nur sehr wenige Studierende (15 Probanden) zuzuordnen.

¹²⁷ Zu den 14 Probanden (1,6 %) von 887 Studierenden liegen keine Informationen bzgl. des Alters vor.

8. Hypothesen: Ergebnisse und Kommentare

In der ersten Alterskategorie (19 bis 25 Jahre) wurde der maximale C-Testwert von 96 Punkten erreicht, gefolgt von der Gruppe der 26- bis 30-jährigen Studierenden mit einem maximalen Wert von 94 Punkten. In den restlichen Altersgruppen ist die maximal erzielte C-Testleistung wesentlich niedriger. Ab dem 40. Lebensjahr wird in der getesteten Gruppe eine starke Abnahme der durchschnittlichen C-Testleistung festgestellt. Zwischen den Kategorien III und IV gibt es einen großen Unterschied von fast 10 Punkten. Zwischen den gebildeten Alterskategorien 1 bis 5 sind die beobachteten Leistungsunterschiede jedoch nicht signifikant, $p = 0.367 > .01$ (Sig. 2-seitig). Das bedeutet, dass die Leistungsunterschiede zwischen den jeweiligen Gruppen zufällig entstanden sein können.

Betrachtet man die Alterskategorien separat, von denen Wockenfuß berichtet, dass hier ein signifikanter Leistungsanstieg zu beobachten sei (vgl. Wockenfuß, 2009, S. 171), so ist in der vorliegenden Studie keine Tendenz zum Leistungsanstieg in Verbindung mit dem Alter zu erkennen (s. Tabelle 28) (vgl. auch Berechnungen im Anhang A, Teil 3.3.3). Vergleicht man die Medianwerte dieser Gruppen miteinander, so liegen diese mit Ausnahme für die Kategorie der 19- und 22-Jährigen Studierenden konstant bei 63 Punkten. Das bedeutet, dass es in der vorliegenden Studie in der Gruppe der Probanden zwischen 19 und 25 Jahren keine nachweisbar vom Alter abhängigen Leistungen gibt.

Tabelle 28: C-Testleistungen der Studierenden zwischen 19 und 25 Jahren

Alter	Häufigkeiten	C-Testleistung (Mittelwert, Median)
19 Jahre	17	64,2 (64)
20 Jahre	165	62,8 (63)
21 Jahre	176	58,9 (63)
22 Jahre	129	59,5 (64)
23 Jahre	92	60,6 (63)
24 Jahre	61	57,4 (63)
25 Jahre	58	55,9 (63)

8. Hypothesen: Ergebnisse und Kommentare

Um zu untersuchen, ob die sprachliche Leistung nach dem Studium ansteigt, wurden zwei Gruppen gebildet: Studierende mit abgeschlossenem Studium und Studierende ohne abgeschlossenes Studium bzw. ohne abgeschlossene Ausbildung (s. Tabelle 28).

Die Ergebnisse aus der Tabelle 29 zeigen eine Tendenz zum Leistungszuwachs bei Personen, die ein abgeschlossenes Studium haben; allerdings gibt es in der dritten Alterskategorie (zwischen 31 und 35 Jahren) eine Leistungsabnahme. Die berechneten Leistungsunterschiede zwischen der Gruppe der Studierenden mit/ohne abgeschlossenem/s Studium sind signifikant, zwischen den einzelnen Altersgruppen jedoch nicht signifikant, $p = 0.281 > .01$ (Sig. 2-seitig).

Tabelle 29: C-Testleistung der Studierenden, aufgeteilt nach Alterskategorien und abgeschlossenem Studium

Alterskategorien	abgeschlossenes Studium	kein abgeschlossenes Studium
I. 19 bis 25 Jahre	61,8 (63)	60,7 (63)
II. 26 bis 30 Jahre	65 (71)	53,6 (57,5)
III. 31 bis 35 Jahre	59,1 (63)	63,4 (74,5)
IV. 36 bis 40 Jahre	76,1 (84)	63,2 (63)
V. älter als 40 Jahre	46,2 (43,5)	-
Gesamt	N = 112	N= 549

Zusammenfassend lässt sich festhalten, dass in der getesteten Gruppe keine altersspezifischen Leistungsunterschiede bestehen, die statistisch signifikant sind. Betrachtet man die Leistungen der Studierenden, die ein Studium abgeschlossen haben, so ergeben sich im Durchschnitt Unterschiede zugunsten der Studierenden mit einem ersten Abschluss. Die Gruppe dieser Studierenden erreicht durchschnittlich höhere C-Testwerte, was vermuten lässt, dass in der Studienzeit ein Leistungszuwachs in verschiedenen sprachlichen Bereichen stattfindet, was sich wiederum auf die sprachliche Leistungen auswirkt, die durch den C-Tests erfasst werden.

Die Hypothese, dass es Unterschiede zwischen den C-Testleistung jüngerer und älterer Probanden innerhalb der Gruppe der getesteten Studierenden bestehen, konnte nur teilweise bestätigt werden: Zwischen den einzelnen Altersgruppen von 19 bis 35 Jahren besteht kein signifikanter Unterschied. Erst die Leistungen von Probanden, die älter sind als 40 Jahre, nehmen signifikant nachweisbar ab.

Ein signifikanter Unterschied lässt sich hingegen zwischen der Gruppe der Probanden feststellen, die bereits eine Ausbildung oder ein Studium abgeschlossen haben und solchen, die sich noch in einem (ersten) Studium befinden. Diese beiden Gruppen sind aber nicht identisch mit einer Aufteilung nach Altersgruppen.

8.1.4 C-Testleistung und Schulabschluss

Hypothese 11: *Zwischen den Testleistungen Studierender mit verschiedenen Schulabschlüssen bestehen Leistungsunterschiede.*

Angenommen wird, dass Absolventen des Gymnasiums höhere sprachliche Leistungen in den C-Tests erzielen als Absolventen anderer Schulformen.

Es wird vermutet, dass Studierende mit einem Gymnasialabschluss höhere Testleistungen erzielen als Studierende mit anderen Schulabschlüssen, da an das Gymnasium Schüler und Schülerinnen empfohlen werden, die bis dahin hohe Schulleistungen gezeigt haben. Falls diese Unterschiede tatsächlich bestehen, sollten sie sich auch in den sprachlichen Leistungen der Studierenden zeigen und mithilfe von C-Tests messbar sein.

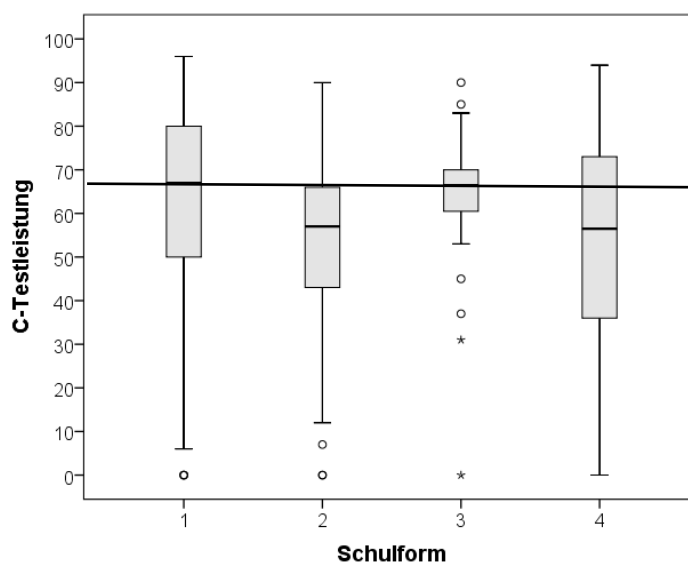
Bei der Frage nach dem erworbenen Schulabschluss wurden den getesteten Studierenden vier Antwortoptionen angeboten (s. den Fragebogen im Anhang A, Teil 1). Mehr als die Hälfte (58,7 %) der Studierenden erwarben ihren Schulabschluss danach an einem Gymnasium, 31,7 % an einer integrierten Gesamtschule und nur 3,2 % an einer Berufsschule¹²⁸.

¹²⁸ 56 Studierende (6,3 %) haben auf die Frage *An welcher Schulform haben Sie Ihren Schulabschluss erworben* die Option *Sonstiges* angekreuzt. Hier kann es sich z. B. um einen

8. Hypothesen: Ergebnisse und Kommentare

In Abb. 16 ist gut zu sehen, dass sich die C-Testleistungen aller vier Gruppen stark voneinander unterscheiden: Studierende mit einem Gymnasialabschluss erreichten im Durchschnitt höhere Testwerte (63,9 Punkte) als Studierende mit anderen Schulabschlüssen (vgl. auch Berechnungen im Anhang A, Teil 3.3.4). Der Medianwert liegt in dieser Gruppe bei 67 Punkten und ist dem Referenznormwert gleich. Die in der Grafik eingezeichnete horizontale Linie dient der besseren Orientierung; sie zeigt den ermittelten Referenznormwert an.

Abb. 16: C-Testleistung und erworbener Schulabschluss



1 - Schulabschluss erworben am Gymnasium; 2 – integrierte Gesamtschule; 3 – Berufsschule, 4 - Sonstiges

Auffällig ist, dass der Medianwert der Gruppe der Studierenden, die ihren Schulabschluss an einer Berufsschule erworben haben, knapp unter dem Medianwert der Gymnasialabsolventen liegt, der 66,5 Punkte beträgt. Die durchschnittliche Leistung dieser Gruppe ist etwas niedriger anzusetzen; sie beträgt 63,5 Punkte. Die erzielten Testleistungen der Studierenden, die ihren Abschluss an einer integrierten Gesamtschule erworben haben, liegen durchschnittlich deutlich unter dem Referenznormwert und betragen nur 54,9 Punkte. Der Median in dieser Gruppe ist etwas höher und erreicht 57 Punkte,

Auslandsabschluss handeln, falls Studierende erst nach ihrer Abiturprüfung nach Deutschland eingewandert sind. 1,2 % (11 Probanden) der Studierenden gaben keine Information bzgl. des Schulabschlusses an.

liegt aber immer noch weit unter dem Referenznormwert. Hier besteht eine Differenz von 10 Punkten.

Einen ähnlichen Durchschnittswert wie bei Studierenden mit Gesamtschulabschluss erreicht die Gruppe der Studierenden, die auf die Frage nach dem erworbenen Schulabschluss *Sonstiges* angekreuzt hat¹²⁹. Die Streuung der Testleistungen ist in dieser Gruppe sehr breit, ähnlich wie in der Gruppe der Probanden mit gymnasialem Schulabschluss, und sie erstreckt sich über alle Leistungsgruppen von 0 bis 94 Punkte.

Am wenigsten streut die Leistung der Studierenden mit einem an einer Berufsschule erworbenen Schulabschluss. Die minimale Leistung beträgt knapp 50 Punkte und die maximale Leistung ca. 83 Punkte. Die auf der Grafik als Kreise und Sternchen gekennzeichneten Testwerte außerhalb des Box-Plots markieren statistische Ausreißer in dieser Gruppe.

Es ist folgende Tendenz festzustellen: Die Leistung der Studierenden mit einem Gymnasialabschluss ist höher als die Leistung der Studierenden mit einem an einer integrierten Gesamtschule erworbenen Schulabschluss. Die Leistung der Studierenden mit einem an einer Berufsschule erworbenen Abschluss ist höher als die Leistung der Studierenden mit einem sonstigen Schulabschluss. Die Leistung der Studierenden mit einem Berufsschulabschluss bestätigt die von Lehmann (2005) untersuchte These, dass Berufsschüler offenbar sprachlich besser qualifiziert sind als vielfach angenommen wird (ULME)¹³⁰. Der Studie zufolge erzielten 82,7% der Berufsschüler zu Beginn ihrer Berufsausbildung ein besseres Leseverständnis im Fach Deutsch als zu Beginn der 9. Klasse. Zwischen den gebildeten Gruppen bestehen signifikante Leistungsunterschiede, $p = .000 < .01$ (Sig. 2-seitig).

¹²⁹ Bei der Gruppe von Studierenden, die auf die Frage nach dem Schulabschluss *Sonstiges* angekreuzt haben, handelt sich um Studierende, die ihren Abschluss in Russland, Polen oder China gemacht haben. In diesen Ländern gibt es Schulabschlüsse, die mit dem deutschen Abitur nicht vergleichbar sind.

¹³⁰ Oliver Schirg, Pressemitteilung vom 05.04.2005, Berufsschüler sind besser als ihr Ruf, abrufbar unter: <http://www.welt.de/print-welt/article563575/Berufsschueler-sind-besser-als-ihr-Ruf.html>, zuletzt gesehen am 06.06.2013.

8. Hypothesen: Ergebnisse und Kommentare

Tabelle 30: C-Testleistung vs. erworbener Schulabschluss

Schulabschluss	Leistungsstufe I	Leistungsstufe II	Leistungsstufe III	Leistungsstufe IV	Leistungsstufe V	Gesamt
Gymnasium	33 6,4 %	80 15,6 %	260 50,6 %	90 10,3 %	51 17,5 %	514 58,6 %
Integrierte Gesamtschule	1 0,4 %	7 2,5 %	156 56,1 %	80 28,8 %	34 12,2 %	278 31,7 %
Berufsschule	1 3,6 %	4 14,3 %	19 67,9 %	2 7,1 %	2 7,1 %	28 3,2 %
Sonstiges	1 1,8 %	6 10,7 %	23 41,1 %	13 23,2 %	13 23,2 %	56 6,4 %
Gesamt	36 4,1 %	97 11,1 %	458 52,3 %	185 21,1 %	100 11,4 %	876 100 %

Legende: Leistungsstufe I: > als 89 Punkte; Leistungsstufe II: 82-89 Punkte; Leistungsstufe III: 53-81 Punkte; Leistungsstufe IV: 36-52 Punkte; Leistungsstufe V: < als 35 Punkte

Teilt man die Leistungen der Studierenden nach dem erworbenen Schulabschluss auf (Tabelle 30), so zeigt sich, dass die Testleistungen der Studierenden mit Gesamtschulabschluss auf den unteren Leistungsstufen (IV und V) gegenüber den anderen Gruppen (z. B. Schulabschluss am Gymnasium und an der Berufsschule) mit 41 % überwiegen. In der Gruppe der Studierenden, die ihren Schulabschluss vermutlich im Ausland erworben haben, liegt die Leistung zu einem ähnlich hohen Anteil (46,4 %) in den untersten Leistungsbereichen IV und V; allerdings kann diese Leistung nicht mit den anderen verglichen werden, da im Ausland erworbene Schulabschlüsse in seltenen Fällen mit in Deutschland erworbenen äquivalent sind. Die höchsten Testleistungen wurden von Absolventen des Gymnasiums erzielt, die auf den oberen Leistungsstufen mit 22 % überrepräsentiert sind.

Zusammenfassend lässt sich festhalten, dass Leistungsunterschiede in der Gruppe der Studierenden mit verschiedenen Schulabschlüssen bestehen. Vergleicht man die C-Testleistungen verschiedener Gruppen miteinander, so zeigen sich signifikante Unterschiede zugunsten der Gymnasialabsolventen, über die auch die großen Leistungsstudien wie PISA (2000) und DESI (2008) berichteten.

Die Hypothese, dass Absolventen des Gymnasiums höhere sprachliche Leistungen in den C-Tests erzielen als Absolventen anderer Schulformen, konnte bestätigt werden.

8.1.5 C-Testleistung und angestrebter Studienabschluss

Hypothese 12: *Zwischen den Studierenden, die verschiedene Qualifikationen im Lehramt anstreben, bestehen Leistungsunterschiede.*

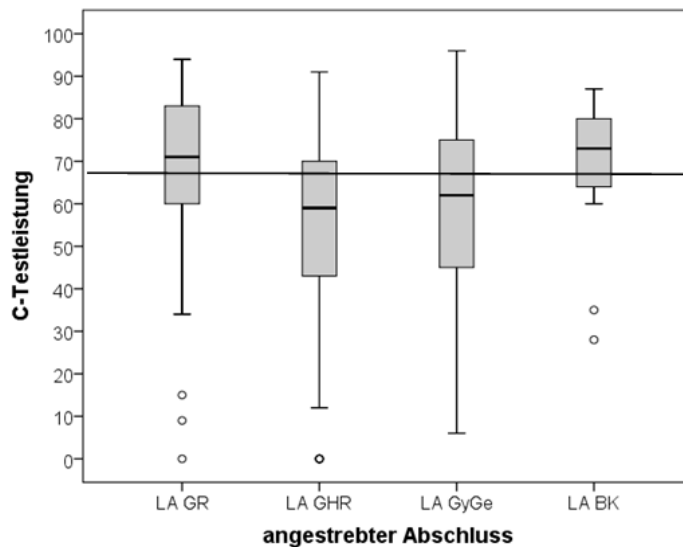
Angenommen wird, dass Studierende mit dem angestrebten Studienabschluss Lehramt Gymnasium/Gesamtschule höhere Testwerte erzielen als Studierende mit anderen Lehramtsoptionen.

Vermutet wird, dass Studierende mit der Lehramtsoption Gymnasium/Gesamtschule (LA GyGe) aufgrund ihrer hohen sprachlichen Selbsteinschätzung und u. a. guter sprachlicher Voraussetzungen für den Lehrerberuf an einem Gymnasium oder einer integrierten Gesamtschule die genannte Lehramtsoption gewählt haben.

Die Zusammensetzung der Studierenden in der vorliegenden Studie unterscheidet sich deutlich von Gruppe zu Gruppe. So ist der Anteil der Studierenden, die den Abschluss Lehramt Gesamt-, Haupt und Realschule (LA GHR) anstreben, mit 34,7 % (308 Probanden) am höchsten, während der Anteil der Studierenden des Lehramts Grundschule (LA Gr) mit 19,5 % (173 Probanden) und des Lehramts Berufskolleg (LA BK) am geringsten ist und nur 3 % (27 Probanden) ausmacht. Die Gruppe der Studierenden mit dem angestrebten Abschluss LA GyGe ist in Bezug auf die Gesamtgruppe verhältnismäßig groß und beträgt 25,9 % (230 Probanden)¹³¹.

¹³¹ 16,8 % der Studierenden (149 Probanden) machten keine Angaben zu Fragen nach dem angestrebten Studienabschluss.

Abb. 17: C-Testleistung vs. angestrebter Abschluss



Aus Abb. 17 ist ersichtlich, dass sich die Leistungen sowohl zwischen den Gruppen als auch innerhalb der Gruppen unterscheiden. Innerhalb der Gruppen zeigt sich eine unterschiedlich breite Varianz (vor allem in der Gruppe der Studierenden mit einem LA GyGe und LA BK anzustrebenden Abschluss).

Durchschnittlich höhere Leistungen erreichen Studierende des Lehramts Grundschule (Mittelwert: 69 Punkte; Median: 71 Punkte) und des Lehramts Berufskolleg (Mittelwert: 70 Punkte; Median: 73 Punkte) (vgl. auch Berechnungen im Anhang A, Teil 3.3.5). Ihre Leistung liegt über dem standardisierten Referenznormwert. Lehramtsstudierende, die den Abschluss Gymnasium/Gesamtschule anstreben, erreichen im Vergleich zu den Studierenden des Lehramts Gesamt-, Haupt und Realschule im Durchschnitt etwas bessere Werte (Mittelwert für GyGe: 60 Punkte; Mittelwert für GHR: 56 Punkte). Der Medianwert der Studierenden des Lehramts Gymnasium/Gesamtschule ist auch etwas höher als der Medianwert der Studierenden des Lehramts GHR (Median für GyGe: 62 Punkte; Median für GHR: 59 Punkte). Die Differenzen der erreichten Testleistungen zwischen Studierenden unterschiedlicher Lehramtsrichtungen sind signifikant, $p = .000 < .01$ (Sig. 2-seitig).

Analysiert man die C-Testleistungen der Studierenden auf unterschiedlichen Leistungsstufen, so ergibt sich folgende Verteilung:

Tabelle 31: C-Testleistungen Studierender verschiedener Lehramtsrichtungen aufgeteilt nach Leistungsstufen (N=738)

Lehramt	Leistungsstufe I	Leistungsstufe II	Leistungsstufe III	Leistungsstufe IV	Leistungsstufe V	Gesamt
Grundschule (Gr)	19 11 %	28 16,2 %	98 56,6 %	24 13,9 %	4 2,3 %	173 23,4 %
Gesamt-, Haupt und Realschule (GHR)	2 0,6 %	15 4,9 %	171 55,5 %	78 25,3 %	42 13,6 %	308 41,7 %
Gymnasium, Gesamtschule (GyGe)	11 4,8 %	27 11,7 %	115 50 %	56 24,3 %	21 9,1 %	230 31,2 %
Berufskolleg (BK)	-	6 22,2 %	19 70,4 %	-	2 7,4 %	27 3,7 %
Gesamt	32 4,3 %	76 10,3 %	403 54,6 %	158 21,4 %	69 9,3 %	738 100 %

Legende: Leistungsstufe I: > als 89 Punkte; Leistungsstufe II: 82-89 Punkte; Leistungsstufe III: 53-81 Punkte; Leistungsstufe IV: 36-52 Punkte; Leistungsstufe V: < als 35 Punkte

Auf den höheren Leistungsstufen I und II sind Studierende des Lehramts Grundschule mit 27,2 % überrepräsentiert. Auf den unteren Leistungsstufen IV und V überwiegen mit 38,9 % die Leistungen Studierender des Lehramts Gesamt-, Haupt und Realschule¹³².

Studierende anderer Gruppen sind auf diesen Leistungsstufen mit kleineren Anteilen repräsentiert, doch immer noch verhältnismäßig groß im Vergleich zu der Gesamtgruppe (LA Gr mit 16,2 %; LA GyGe mit 33,4 %; LA BK mit 7,4 %). Die Leistungsunterschiede zwischen den gebildeten Gruppen sind signifikant. Dies macht die Notwendigkeit der Förderung sprachlicher Fähigkeiten in allen Gruppen deutlich.

¹³² Eine mögliche Erklärung hierfür könnten unterschiedliche Studienanforderungen in den Fächern sein (vgl. Studienordnungen an der Universität DuE für das Lehramt Gr, GHR, GyGe und BK), abrufbar unter http://www.uni-due.de/germanistik/la_lpo2003.shtml, zuletzt gesehen am 11.06.2013. Das Studium mit der Lehramtsoption GHR wird beispielsweise von Studienbewerbern im Vergleich zu den anderen (z. B. LA GyGe) als ein leichter Studiengang angesehen, da weniger Module mit weniger Semesterwochenstunden sowohl im Grund- als auch im Hauptstudium absolviert werden müssen, was die Studienzeiten verkürzen kann. Außerdem wird im LA GHR kein Latinum verlangt, was von einigen Studierenden als zusätzliche Hürde angesehen wird. Daher wird das Lehramt GHR vermutlich von leistungsschwächeren Studienbewerbern gewählt, deren sprachliche Leistungen laut C-Testergebnissen eher unteren Leistungsstufen zuzuordnen sind (s. Tabelle 8.13). Auch das Lehramt Grundschule wird von leistungsschwächeren Studierenden als schwieriger eingeschätzt als das Lehramt GHR, weil beim Grundschulstudium obligatorisch das Fach Mathematik absolviert werden muss.

Zusammenfassend lässt sich festhalten, dass zwischen den Studierenden, die verschiedene Qualifikationen im Lehramt anstreben, signifikante Leistungsunterschiede bestehen. Die Annahme, dass Studierende mit dem angestrebten Abschluss Lehramt Gymnasium/Gesamtschule höhere C-Testleistungen erzielen als Studierende anderer Lehramtsoptionen, hat sich nicht bestätigt. In der vorliegenden Untersuchung erreichen Studierende des Lehramts Grundschule, die in der getesteten Gesamtgruppe mit 19,5 % vertreten sind, im Durchschnitt signifikant bessere C-Testleistungen. Obwohl Studierende des Lehramtes Berufskolleg in der vorliegenden Gruppe durchschnittlich höhere Leistungen erzielen als Studierende anderer Gruppen, soll dieser Befund nicht überinterpretiert werden, denn die Gruppe der Studierenden mit der Lehramtsoption BK enthält nur 3 % (27 Probanden).

Die Hypothese, dass Studierende mit dem angestrebten Studienabschluss Lehramt Gymnasium/Gesamtschule höhere Testwerte erzielen als Studierende mit anderen Lehramtsoptionen, konnte nicht bestätigt werden. Die höchsten Werte werden von Studierenden des Lehramts für die Grundschule und für das Berufskolleg erzielt.

8.2 Hypothesen zur Validität von C-Tests

Bevor weitere Hypothesen vorgestellt werden, sollen im Vorfeld verschiedene Validitätsarten vorgestellt werden, die zum Verständnis der Interpretation der berechneten Zusammenhänge beitragen¹³³.

Wie bereits in Kap. 7 beschrieben, ist Validität eines der wichtigsten Testgütekriterien. Valide ist ein Test dann, wenn er misst, was er zu messen vorgibt. Bei einer Höhe von $r = 0.30$ bis 0.65 werden Validitätswerte als ausreichend akzeptiert (Bortz & Döring, 2006). Allerdings muss die Höhe des Zusammenhangs in Abhängigkeit vom zu messenden Merkmal und dem zugrunde liegenden Testkonstrukt interpretiert werden. Es werden verschiedene

¹³³ Ausführliche Informationen zum Begriff Validität und Validitätsbestimmung in Kap. 7.

Validitätsarten unterschieden: Kriteriumsvalidität (kriterienbezogene Validität) und Konstruktvalidität¹³⁴. Zur Beschreibung weiterer Hypothesen, die durch korrelationsstatistische Untersuchungen mit verschiedenen Außenkriterien überprüft werden, kommen zwei Validitätsarten in Frage, die der kriterienbezogenen Validität zuzuordnen sind: die *Vorhersagevalidität* und die *Übereinstimmungsvalidität* (vgl. ALTE, Modul 1, S.11; Rey, 2012, S.64).

Vorhersagevalidität (prognostische Validität): Die Testleistung lässt Vorhersagen auf spätere Ereignisse zu, z. B., ob ein Studienbewerber mit gutem Eingangstestwert in einem spezifischen Test für Mediziner auch ein guter Medizinstudent sein wird. Analysen in Bezug auf die prognostische Validität von C-Tests als Lesetests befinden sich in Kap. 7.3.

Übereinstimmungsvalidität (konkurrente Validität): Das Testergebnis stimmt mit einem Außenkriterium überein. Man könnte z. B. überprüfen, ob ein Schüler mit einem hohen IQ-Testwert auch von dem Lehrer als intelligent eingestuft wird.

8.2.1 C-Testleistung und Deutschnote

Hypothese 13: Studierende mit einer guten Abiturnote im Fach Deutsch erzielen im Durchschnitt signifikant bessere C-Testleistungen als Studierende mit einer darunter liegenden Deutschnote.

Studien zur Übereinstimmungsvalidität von C-Testleistungen und Deutschnoten in verschiedenen Klassenstufen haben ergeben, dass bei jüngeren Schülern ein signifikant höherer Zusammenhang besteht als bei Studierenden. Gummich (1997)¹³⁵ kam zu dem Ergebnis, dass mit zunehmender Klassenstufe in der Sek. I die Übereinstimmungsvalidität sinkt, was später auch bei Wockenfuß (2009) belegt wurde.

Baur & Spettmann (2005) bestätigten die Überschneidungen der C-Testergebnisse mit den im Fach Deutsch vergebenen Zeugnisnoten im Fremdsprachenbereich. Untersucht wurden Fremdsprachenkenntnisse in Deutsch bei griechischen Schülern der 10. Jahrgangsstufe im Alter von 15-16

¹³⁴ Näheres zur Konstruktvalidität in Kap. 7.

¹³⁵ Unveröffentlichte Arbeit.

8. Hypothesen: Ergebnisse und Kommentare

Jahren und bei Schülern der 12. Jahrgangsstufe im Alter von 17 Jahren. In der Studie konnte belegt werden, dass in beiden Gruppen die durchschnittliche Lösungsquote der Schüler mit der Note 2 über der durchschnittlichen Lösungsquote der Schüler mit der Note 3 und diese wiederum über der von Schülern mit der Note 4 liegt¹³⁶. Später konnten die Autoren ebenfalls kleine bis mittlere Zusammenhänge bei muttersprachlichen C-Tests für die Klassenstufen 4-6 belegen (vgl. Baur & Spettmann, 2006a,b, 2007, 2008).

Bei der Überprüfung der Validität von muttersprachlichen C-Tests für ältere Klassenstufen stellte Wockenfuß (2009) fest, dass die Korrelation zwischen der C-Testleistung und der Deutschzeugnisnote mit zunehmendem Alter geringer ist. Allerdings betrifft das nicht die Abiturnote; hier wurden höhere Validitätskoeffizienten beobachtet ($-0.37 \geq r_{(tc)} \geq -0.44$). Der höchste signifikante Zusammenhang zwischen der C-Testleistung und der Deutschnote wurde bei den jüngeren (9- und 10-jährigen) Probanden (N=107) festgestellt und betrug $r_{(tc)} = -.51$; in den darauffolgenden Altersgruppen bis einschließlich der 16-jährigen streuten die Koeffizienten in engen Grenzen ($-0.25 \geq r_{(tc)} \geq -0.33$) bis auf eine Ausnahme bei den 15-Jährigen.

Wockenfuß betont, dass die Annahme, dass Deutschnoten sprachliche Fähigkeiten erfassen, überdacht werden muss. Sie kommt in ihrer Studie zu dem Schluss,

„... dass muttersprachliche C-Tests am Anfang der Sekundarstufe I eher unterrichtsrelevante sprachliche Fähigkeiten messen, die auch der Lehrer in seiner Note beurteilt, [sic!] jedoch im Verlauf der Sekundarstufe I und in der Oberstufe jedoch nicht mehr [...]“ (Wockenfuß, 2009, S. 175).

Im *Sprachkompetenzprojekt Diagnose und Förderung* wurde zwischen der C-Testleistung und der Abiturnote im Fach Deutsch ein signifikanter Zusammenhang von Pearsons $r = -0.257$, ($p = .000 < .01$) festgestellt. Getestet wurden 1.842 Lehramtsstudierende im Grundstudium¹³⁷.

¹³⁶ www.daf-netzwerk.org/tagungen/ergebnisse/.../Tests-GER-DaF.doc, zuletzt gesehen am 06.03.2013.

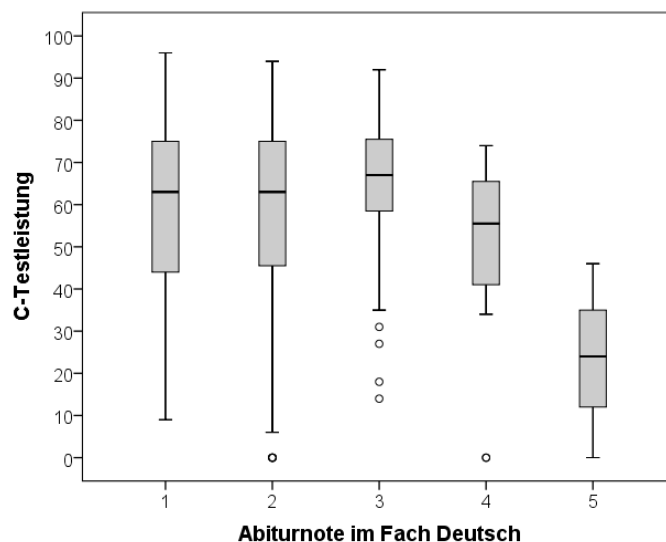
¹³⁷ Die Daten wurden im WS 2012 berechnet und dem Vortrag im TestDaF-Institut entnommen.

8. Hypothesen: Ergebnisse und Kommentare

Der Berechnung der Korrelationen in der vorliegenden Studie standen die Daten von 659 Studierenden zur Verfügung¹³⁸. Ein Viertel der Studierenden, 40,9 % (363 Probanden), gaben an, dass sie im Fach Deutsch eine Zwei haben, 17,7 % eine Eins (157 Probanden), 12,6 % (112 Probanden) haben im Fach Deutsch eine Drei und nur 2,7 % (24 Probanden) eine Vier. Insgesamt gibt es nur 0,3 % Studierende (3 Probanden) mit der Abiturnote Fünf im Fach Deutsch.

Die unten aufgeführte Abbildung zeigt die Streuung innerhalb der einzelnen Gruppen. Sichtbar ist, dass die Streuungen unterschiedlich groß sind (z. B. Gruppe der Probanden mit der Deutschnote Vier).

Abb. 18: C-Testleistung vs. Abiturnote im Fach Deutsch



Vergleicht man die durchschnittlichen C-Testleistungen der Studierenden miteinander, so ergeben sich signifikante Leistungsunterschiede zwischen den Studierenden mit besseren und schlechteren Noten (vgl. dazu Berechnungen im Anhang A, Teil 3.2.1). Die Differenz zwischen der Abiturnote *ausreichend* und der Note *mangelhaft* spiegelt sich auch deutlich in den C-Testergebnissen wider. Während Studierende mit der Deutschnote Ausreichend im Durchschnitt ebenfalls signifikant schlechter abschneiden als die Studierenden mit befriedigenden, guten oder sehr guten Deutschnoten, verwischt sich dieser Unterschied zwischen den drei Gruppen mit sehr guter, guter und befriedigender

¹³⁸ In der Befragung nannten 228 Studierende (25,7 %) von 887 keine Deutschnote. Viele Studierende gaben als Grund an, dass sie sich an die Abiturnote nicht mehr erinnern, andere wiederum, dass sie aus Gründen der Anonymität die Noten nicht preisgeben wollten.

8. Hypothesen: Ergebnisse und Kommentare

Deutschnote im Abitur. Hier sagt die Deutschnote das Abschneiden im C-Test nicht mehr voraus. Die Studierenden mit der Abiturnote *ausreichend* liegen sogar geringfügig über den durchschnittlichen Leistungen der Studierenden mit der Abiturnote sehr gut und gut.

Berechnet man die Korrelationen zwischen den C-Testleistungen und den Abiturnoten in Deutsch, so ergibt sich ein schwacher Zusammenhang von Pearson $r = -0.028$, der für die vorliegenden Daten nicht signifikant ist, $p = 0.478 > .005$ (Sig. 2-seitig). Das bedeutet, dass zwischen der Abiturnote im Fach Deutsch und der C-Testleistung in der vorliegenden Untersuchung ein schwacher Zusammenhang besteht.

Vergleicht man die auf der Grafik gekennzeichneten Quartile, so sieht man, dass Studierende mit der Abiturnote Drei teilweise in den unteren Leistungsbereichen höhere C-Testergebnisse erzielen als Studierende mit besseren Deutschnoten (1 oder 2). Vermutlich spielen hier zwei Faktoren eine Rolle: Zum einen werden die Halbjahresnoten, die in die Berechnung der Abiturnote einfließen, von den Lehrpersonen subjektiv vergeben, und zum anderen werden im Fach Deutsch in der Oberstufe nicht nur sprachliche Fähigkeiten wie Grammatik, Orthografie, Semantik oder Lexik bewertet, die in den niedrigeren Klassenstufen stärker in die Leistungsbewertung einfließen. Dies würde die Hypothese von Wockenfuß stützen, wonach Noten in den höheren Altersstufen nicht mehr die Einschätzung der sprachlichen Fähigkeiten widerspiegeln.

Resümierend lässt sich festhalten, dass die Deutschnote das individuelle C-Testergebnis nicht zuverlässig vorhersagen kann, da die Varianzen in den Gruppen sehr groß sind. Das bedeutet, dass z. B. mehr als die Hälfte der Abiturienten mit einem Befriedigend im Fach Deutsch bessere Leistungen im C-Test zeigen als Abiturienten mit der Note sehr gut. Eine Vergrößerung der Stichprobe hat den Korrelationskoeffizient stabilisiert, wie die Befunde aus dem Sprachkompetenzprojekt zeigen (s.o.). Bei weiteren Untersuchungen könnte die Frage nach der Abiturnote aus dem Fragekatalog herausgenommen werden, da vermutlich mit der Abiturnote – jedenfalls in den höheren Klassenstufen – nicht nur sprachliche Fähigkeiten der Schüler erfasst werden, auf die sich die hier vorgestellte Untersuchung fokussiert.

Diese Hypothese, dass Studierende mit einer guten Abiturnote im Fach Deutsch im Durchschnitt signifikant bessere C-Testleistungen erzielen als Studierende mit einer darunter liegenden Deutschnote, wurde bestätigt.

8.2.2 C-Testleistung und Englischnote

Hypothese 14: *Studierende mit einer guten Abiturnote im Fach Englisch erzielen im Durchschnitt signifikant bessere C-Testleistungen als Studierende mit einer darunter liegenden Englischnote.*

Angenommen wird, dass eine intensive Beschäftigung mit der Fremdsprache Englisch verschiedene Dimensionen der Sprache (phonologisch, syntaktische, semantische usw.) aktiviert und weiterentwickelt. Eine hohe fremdsprachliche Kompetenz könnte sich somit positiv auf die erfolgreiche Rekonstruktion der C-Tests auswirken.

In der vorliegenden Studie gibt es eine breite Streuung der Abiturnoten im Fach Englisch. Nach den erhobenen Daten erhielten 28,2 % der Studierenden (250 Probanden) als Abiturnote eine Zwei und 29,3 % eine Drei (260 Probanden). Nur 4,7 % (42 Probanden) der Studierenden haben die Schule mit einer Eins im Fach Englisch absolviert. 8,1 % der Studierenden haben im Abiturzeugnis im Fach Englisch eine Vier (72 Probanden) und 0,9 % (8 Probanden) eine Fünf. Insgesamt wurden die Daten von 632 Probanden ausgewertet.¹³⁹

Die Abbildung 19 zeigt, dass die Streuungen innerhalb der Gruppe sehr unterschiedlich sind. Die stärkste Streuung weist die Gruppe der Probanden mit der Note Drei im Fach Englisch auf. Wenn man die Medianwerte aller Gruppen zu einer geraden Linie verbindet, so zeichnet sich eine steigende Tendenz zwischen den Medianwerten der Gruppe mit einer Eins und einer Fünf ab. Eine Ausnahme bildet hier die Gruppe der Probanden mit der Englischnote Vier; der Medianwert liegt hier über der gekennzeichneten Linie. Das bedeutet, dass im Durchschnitt eine bessere Englischnote einer besseren C-Testleistung und eine schlechtere

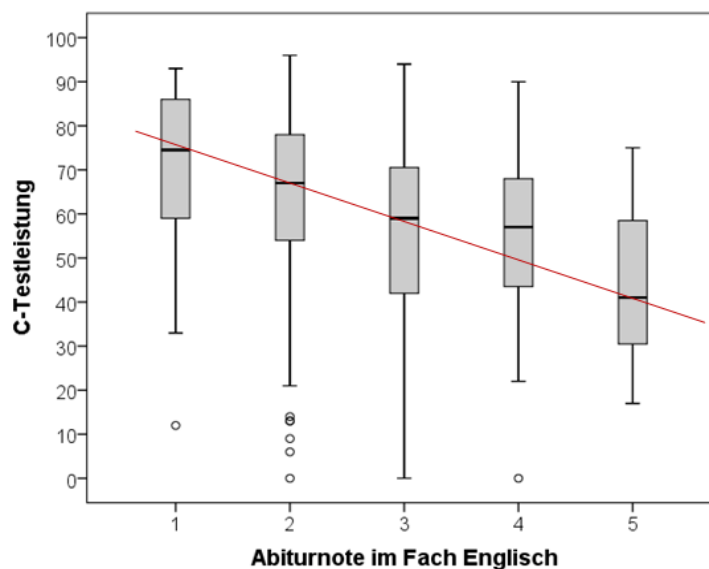
¹³⁹ 28,7 % der Studierenden (255 Probanden) gaben bei der Befragung keine Informationen über die Abiturnote im Fach Englisch. Ein möglicher Grund könnte sein, dass nicht alle Studierenden in der Schule Englisch als Fach hatten.

8. Hypothesen: Ergebnisse und Kommentare

Englischnote einer schwächeren C-Testleistung entspricht. Jedoch trifft eine solche Zuordnung nicht bei allen Studierenden zu, d. h. es kann vorkommen, dass eine überdurchschnittliche C-Testleistung auch von einem Studierenden erreicht werden kann, wenn er eine schlechte Englischnote im Abiturzeugnis hat (vgl. den letzten Box-Plot in Abb. 19).

Es wurde ein geringer signifikanter Zusammenhang zwischen der Englischnote und dem C-Testergebnis festgestellt, $r = -0.240$ ($N = 632$), $p = .000 < .01$ (Sig. 2-seitig) (vgl. dazu Berechnungen im Anhang A, Teil 3.2.2).

Abb. 19: C-Testleistung vs. Abiturnote im Fach Englisch



Der kleine Korrelationswert deutet darauf hin, dass die Englischnote kein zuverlässiger Prädiktor für die Einschätzung der sprachlichen Leistungen im C-Test ist. Auch wenn das Fach Englisch eine intensive Auseinandersetzung mit verschiedenen sprachlichen Komponenten vorsieht (z. B. im phonologischen, syntaktischen, semantischen Bereichen usw.), kann nicht darauf geschlossen werden, dass dies Auswirkungen auf die Steigerung der Kompetenz in der Muttersprache hat.

Zusammenfassend lässt sich festhalten, dass zwischen den durchschnittlichen C-Testleistungen der Studierenden mit guter bzw. sehr guter Englischnote und Studierenden mit einer schlechten bzw. sehr schlechten Englischnote signifikante Leistungsunterschiede bestehen (vgl. dazu Berechnungen im Anhang A, Teil 3,

Tabellen 54-63). Obwohl der Zusammenhang sich bestätigen lässt, sind die Überschneidungen der Gruppen (vgl. oben das Resümee zum Zusammenhang von Abiturnote Deutsch und C-Testleistung) immer noch so groß, dass eine Voreinschätzung der C-Testleistungen im Deutschen auf der Basis der Englischnote unmöglich ist. Etwa die Hälfte der Abiturienten mit einem *Ausreichend* in Englisch erreichen exakt vergleichbare Ergebnisse wie Dreiviertel der Gruppe der Abiturienten mit einem *Sehr gut* in Englisch. Deshalb kann diese Frage aus dem Fragenkatalog ebenso herausgenommen werden.

Die Hypothese, dass Studierende mit einer guten Abiturnote im Fach Englisch im Durchschnitt signifikant bessere C-Testleistungen erzielen als Studierende mit einer darunter liegenden Englischnote, wurde bestätigt.

8.2.3 C-Testleistung und Selbsteinschätzung der Sprachkompetenz

Hypothese 15: *Studierende, die ihre Sprachkompetenz im Deutschen als sehr gut oder gut einschätzen, erzielen auch bessere C-Testleistungen als Studierende, die ihre Sprachkompetenz schlecht einschätzen.*

Angenommen wird, dass die Variable ‚Selbsteinschätzung der Sprachkompetenz‘ als ein aussagekräftiger, zusätzlicher Prädiktor für die C-Testleistungen in der Testung genutzt werden kann.

Erwartet wird, dass sich die von den Studierenden eingeschätzten sprachlichen Fähigkeiten in den C-Testleistungen widerspiegeln, z. B. als gut oder sehr gut eingeschätzte Sprachkompetenz würde wahrscheinlich einem C-Testwert entsprechen, der höheren Leistungsstufen zugeordnet ist.

Lehramtsstudierende sollen aufgrund der Spezifik ihres Berufs in der Zukunft Leistungen von Schülern und Schülerinnen bewerten. Sie sollten dazu in der Lage sein, auch ihre eigenen sprachlichen Fähigkeiten im Deutschen objektiv bewerten zu können. Bei der Frage nach der Selbsteinschätzung der Sprachkompetenz im Deutschen, die den Studierenden in der Testung vorgelegt wurde, wurde nicht spezifiziert, was als Kriterium oder Deskriptor für die Einschätzung von Sprachkompetenz gelten soll. Studierende sollten mithilfe

8. Hypothesen: Ergebnisse und Kommentare

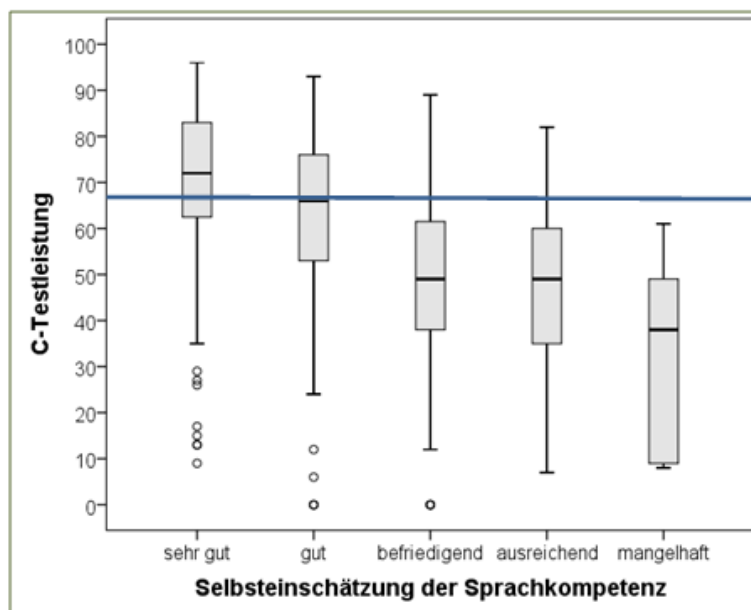
einer fünfstufigen Skala, die dem schulischen Notensystem angeglichen war, ihre allgemeine Sprachkompetenz von *sehr gut* bis *mangelhaft* global einschätzen.

Angenommen wird, dass Studierende ihre sprachlichen Leistungen objektiv einschätzen können und diese Einschätzungen Übereinstimmungen mit der Bewertung der sprachlichen Leistungen durch die eingesetzten C-Tests zeigen.

Im *Sprachkompetenzprojekt Diagnose und Förderung* wurde eine ähnliche Frage an Lehramtsstudierende (N = 1139) gestellt. Festgestellt wurde ein kleiner signifikanter Zusammenhang von Pearsons $r = -0.364$, ($p = .000 < .001$, Sig. 2-seitig).

In der vorliegenden Untersuchung hingegen korrelieren die C-Testleistungen mit den Selbsteinschätzungen der Studierenden etwas stärker. In der Gesamtgruppe ergab sich ein Korrelationskoeffizient von Pearsons $r = -0.430$. Die Korrelation ist signifikant, ($p = .000 < .001$, Sig. 2-seitig). Auf die Frage nach der Einschätzung der Sprachkompetenz antworteten 76,2 % der Studierenden (676 Probanden)¹⁴⁰. In der Abbildung 50 verdeutlichen die Varianzen in den einzelnen Gruppen, ob Studierende im Vergleich zu den C-Testergebnissen ihre sprachlichen Fähigkeiten über- oder unterschätzt haben.

Abb. 20 C-Testleistung vs. globale Selbsteinschätzung der Sprachkompetenz im Deutschen



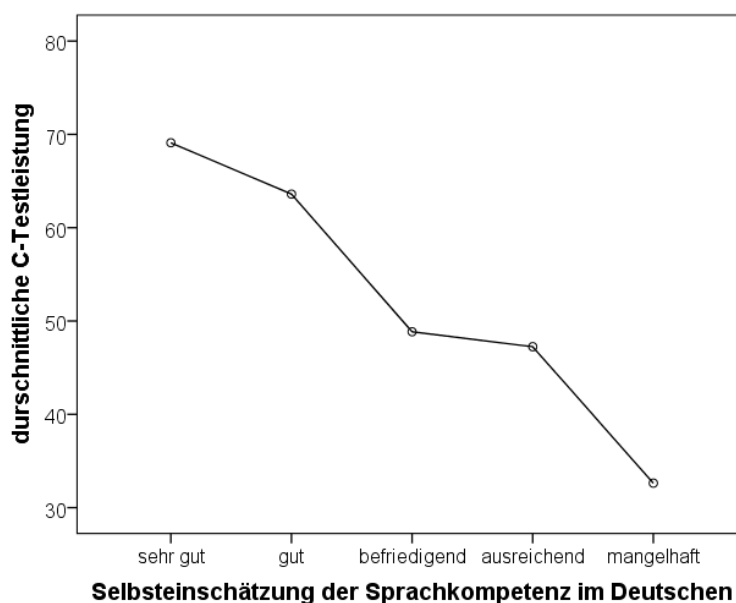
¹⁴⁰ 23,8 % der Studierenden (211 Probanden) gaben auf die Frage keine Antwort.

8. Hypothesen: Ergebnisse und Kommentare

Vergleicht man die C-Testergebnisse mit dem standardisierten Referenznormwert (in der Grafik als horizontale Linie gekennzeichnet), so ergibt sich, dass die durchschnittliche C-Testleistung der Studierenden, die ihre Sprachkompetenz als *sehr gut* eingeschätzten (N = 163; 18,4 %), über dem Gruppennormwert liegt. Ca. 50 % der Studierenden, die ihre Sprachkompetenz im Deutschen als *gut* bewerteten (N = 301; 33,9 %), erreichten hohe oder sehr hohe C-Testwerte. Die Testleistungen in den Gruppen mit der Selbsteinschätzung *befriedigend* (N = 155; 17,5 %) und *ausreichend* (N = 49; 5,5 %) liegen zu ca. 75 % unter dem Gruppennormwert. Die Testergebnisse in der Gruppe der Studierenden, die ihre sprachlichen Fähigkeiten als *mangelhaft* einschätzten (N = 8; 0,9 %), liegen deutlich unter dem Gruppennormwert. Die höchste Leistung beträgt in dieser Gruppe nur knapp 60 Punkte.

Lässt man nur die durchschnittliche C-Testleistungen der oben beschriebenen Gruppen der Studierenden anzeigen (vgl. Abb. 21), so ist eine ansteigende Tendenz festzustellen: Je besser die Einschätzung der Sprachkompetenz der Studierenden, umso höher die erzielte C-Testleistung. Die Leistungsunterschiede zwischen den Gruppen sind signifikant (vgl. Berechnungen im Anhang A, Teil 3, Tabelle 64-73).

Abb. 21: Durchschnittliche C-Testleistungen vs. Selbsteinschätzung der Sprachkompetenz im Deutschen



8. Hypothesen: Ergebnisse und Kommentare

Trotz der bekannten Problematik der Selbsteinschätzung in Form von Über- und Unterschätzung kann diese Frage beim Einsatz von C-Tests als zusätzliche Information zum Testergebnis verwendet werden, da sie signifikante mittlere Korrelationen mit den in den C-Tests erbrachten Leistungen liefert.

Vermutlich würde die Übereinstimmung zwischen Selbsteinschätzung und C-Testleistung steigen, wenn man die Sprachkomponenten, die die Studierenden einschätzen sollen, spezifizierte, wie es z. B. bei den Selbsteinschätzungsskalen im GER der Fall ist (vgl. Items zu Selbsteinschätzung von DIALANG, Europarat, 2001, S. 223ff.).

Zu betonen ist allerdings, dass es eine *zusätzliche* Information ist, die insbesondere dann von Bedeutung ist, wenn die Selbsteinschätzung deutlich über dem tatsächlichen Sprachniveau liegt. Man kann davon ausgehen, dass in diesem Fall die Bereitschaft, sich sprachlich fördern zu lassen, dem betroffenen Studierenden schwieriger zu vermitteln ist.

Resümierend kann festgehalten werden, dass zwischen den C-Testleistungen der Studierenden, die ihre Sprachkompetenz im Deutschen als sehr gut oder gut einschätzen und Leistungen der Studierenden, die ihre Sprachkompetenz im Deutschen als schlecht oder sehr schlecht einschätzen, signifikante Leistungsunterschiede bestehen. Studierende, die ihre Sprachkompetenz gut einschätzen, erzielen signifikant bessere C-Testleistungen als Studierende, die ihre Sprachkompetenz als schlecht einschätzen.

Die Frage nach der *Selbsteinschätzung der Sprachkompetenz* sollte auch in den zukünftigen Erhebungen Bestandteil des Fragenkatalogs bleiben, da sie zusätzliche nützliche Informationen in Bezug auf die Bewertung der sprachlichen Kompetenzen der Studierenden liefert. Diese können u. a. bei der individuellen Beratung von Studierenden genutzt werden. Der Zusammenhang zwischen C-Testleistung und Selbsteinschätzung ist größer als der Zusammenhang zwischen Deutschnote und Englischnote. Die Fähigkeit der Studierenden zur zutreffenden Selbsteinschätzung ist also gut entwickelt – ein Ansatzpunkt auch für die Konzeption von Fördermaßnahmen.

Die Hypothese, dass Studierende, die ihre Sprachkompetenz im Deutschen als sehr gut oder gut einschätzen, im Durchschnitt bessere C-Testleistungen erzielen als Studierende, die ihre Sprachkompetenz schlecht einschätzen, wurde bestätigt.

Die im Weiteren formulierten Hypothesen beziehen sich auf die zusätzlich durch den Fragebogen erfassten Informationen zur Selbsteinschätzung der Lesekompetenz und der schriftlichen Kompetenz. An dieser Stelle soll darauf hingewiesen werden, dass diese Hypothesen zusätzlich aufgestellt werden; sie bilden nicht den Kern dieser Untersuchung.

8.2.4 C-Testleistung und Selbsteinschätzung der Lesekompetenz

Hypothese 16: *Studierende, die ihre Lesekompetenz im Deutschen gut oder sehr gut einschätzen, erzielen höhere C-Testleistungen als Studierende, die ihre Lesekompetenz im Deutsch schlecht einschätzen.*

Angenommen wird, dass sich die von den Studierenden selbst eingeschätzten Lesefähigkeiten durch die in den C-Test ermittelten Leistungen abbilden lassen. Wird die Lesekompetenz als gut bis sehr gut (oder befriedigend) eingeschätzt, so wird ein Testwert erwartet, der den höheren bzw. (niedrigeren) Leistungsstufen zuzuordnen ist.

Insgesamt korrelieren C-Tests mit der Teilkompetenz Leseverstehen. In mehreren Studien wurden mittlere bis hohe Korrelationen festgestellt (vgl. Babaii & Ansary, 2001; Arras et al., 2002; Eckes & Grotjahn, 2006). Bis jetzt gab es keine Untersuchungen für muttersprachliche C-Tests in Kombination mit den Lesetests und der Selbsteinschätzung der Studierenden in Bezug auf ihre Lesekompetenz. Aufgrund der festgestellten Korrelationen zwischen den C-Tests und den Lesetests in der vorliegenden Studie (vgl. Kap. 7) wird davon ausgegangen, dass die selbst eingeschätzten Leseleistungen mit den Leistungen, die sich durch die C-Tests messen lassen, korrelieren.

8. Hypothesen: Ergebnisse und Kommentare

Bei erwachsenen Muttersprachlern werden Lesefähigkeiten selten erhoben, da davon ausgegangen wird, dass spätestens mit dem schulischen Abschluss die Lesekompetenz soweit entwickelt ist, dass sie für ein erfolgreiches Studium, eine berufliche Ausbildung oder für den ausgeübten Beruf ausreicht im Sinne der in der PISA-Studie (2000) formulierten Definition von Lesekompetenz. Untersucht wurden bis jetzt Lesefähigkeiten von jüngeren Erwachsenen (15-jährige Schüler) z. B. in Studien wie PISA (2000, 2003, 2009) oder DESI (2008). Weitere Studien, die Lesefähigkeiten von erwachsenen Sprechern erheben, beziehen sich auf den Fremdsprachenbereich und überprüfen die Teilkompetenz Lesen im Rahmen von Sprachprüfungen oder Sprachtests (vgl. dazu Kap. 6).

In der vorliegenden Untersuchung wurde bei der Frage nach der Selbsteinschätzung der Lesekompetenz den Studierenden eine fünfstufige Skala vorgelegt. Studierende sollten ihre Lesekompetenz global von *sehr gut* bis *mangelhaft* einschätzen; die Teilkomponenten der Lesekompetenz wurden nicht spezifiziert. Die Frage nach der Einschätzung der Lesekompetenz wurde erst im Endstadium der Erhebung (im Jahr 2012) in den Fragenkatalog aufgenommen. Daher liegen nicht zu allen Probanden, die im Laufe der Studie an der Testung teilgenommen haben, Daten vor. Insgesamt wurden Daten von 116 Studierenden erhoben. Bei der Überprüfung des Zusammenhangs wurde in der Gesamtgruppe der Studierenden ein mittlerer Zusammenhang zwischen der C-Testleistung und der selbst eingeschätzten Lesekompetenz beobachtet. Die berechnete Korrelation nach Pearson beträgt $r = -0.404$ und ist auf einem Niveau von 0.01 (2-seitig) signifikant. Das bedeutet, dass die Selbsteinschätzung der Lesefähigkeiten mit den C-Testergebnissen, die fünf verschiedenen Leistungsstufen zugeordnet werden, mittlere Übereinstimmungen hat. Schätzen die Studierenden ihre Lesekompetenz als befriedigend ein, kann es vorkommen, dass die C-Leistung nicht nur der mittleren Leistungsstufe zugeordnet werden kann, sondern häufiger auch einer niedrigeren bzw. höheren Stufe. Bei einem mittleren Zusammenhang können solche Abweichungen häufiger erwartet werden. Da einige Studierende in der untersuchten Gruppe einen Migrationshintergrund haben (vgl. Kap. 8.1.2), erschien es sinnvoll, zwei Gruppen zu bilden: die Gruppe der Studierenden mit und ohne Migrationshintergrund. Für beide Gruppen

8. Hypothesen: Ergebnisse und Kommentare

werden die Korrelationen separat berechnet. Die Analysen zeigen, dass in der Gruppe der Studierenden mit Migrationshintergrund ($N = 54$) im Vergleich zu der Gesamtgruppe ein stärkerer Zusammenhang festzustellen ist. Die berechnete Korrelation nach Pearson beträgt $r = -0.520$, $p = .000 < .01$ (Sig. 2-seitig) und ist als mittlerer Zusammenhang zu interpretieren. Der Anstieg kann durch eine größere Varianz in der Gruppe der Studierenden mit Migrationshintergrund erklärt werden, der z. B. in der Gruppe der Studierenden ohne Migrationshintergrund nicht gegeben ist.

Bei der Berechnung des Zusammenhangs in der Gruppe der Studierenden ohne Migrationshintergrund ($N = 59$) fällt der Korrelationskoeffizient niedriger aus als in der Gruppe der Studierenden mit Migrationshintergrund. Der Korrelationskoeffizient beträgt $r = -0.253$, $p > .05$ (Sig. 2-seitig); damit ist er nicht signifikant und als sehr gering einzuschätzen. Das würde bedeuten, dass zwischen der Einschätzung der mündlichen Kompetenz bei Studierenden mit Migrationshintergrund ein stärkerer Zusammenhang besteht als bei Studierenden ohne Migrationshintergrund. Allerdings muss dieser Befund mit Vorsicht interpretiert werden, da der Stichprobenumfang für die untersuchten Variablen zu klein ist, um Schlussfolgerungen zuzulassen.

Zusammenfassend lässt sich festhalten, dass die durch die muttersprachlichen C-Tests miterfassten Lesefähigkeiten der Studierenden mit der Selbsteinschätzung der Lesekompetenz im Deutschen in Bezug auf die Gesamtgruppe der Studierenden nicht korrelieren. Untersucht man die Gruppe getrennt nach deutschsprachigen Muttersprachlern und Studierenden mit Migrationshintergrund, so zeigt sich ein Unterschied. Diese Befunde sollten jedoch noch mit anderen Testformaten zur Messung der Lesefähigkeiten überprüft werden. Die vorliegenden Daten beruhen nur auf Selbsteinschätzung der Studierenden, und nur in der Gruppe der Studierenden mit Migrationshintergrund bestehen mittlere Zusammenhänge. In der Gruppe der Studierenden mit Muttersprache Deutsch sind sie jedoch sehr niedrig und nicht signifikant (vgl. dazu Berechnungen im Anhang A, Teil 3.2.4). Dies bedeutet, dass Studierende mit Migrationshintergrund ihre Lesekompetenz in Bezug auf die durch die C-Tests miterfassten Lesefähigkeiten besser einschätzen.

Die Frage nach der *Selbsteinschätzung der Lesekompetenz* von Studierenden könnte in zukünftigen Testungen in den Fragekatalog aufgenommen werden, da sie zusätzliche Informationen in Bezug auf die Selbstbewertung der Studierenden liefert und nach der Testung im Rahmen der universitären Beratung genutzt werden kann. Schätzen die Studierenden ihre Lesekompetenz als mangelhaft oder befriedigend ein, erscheint es für die Beratenden einfacher, auf die Bedürfnisse der Studierenden einzugehen und ein passendes Förderangebot im Bereich Lesen anzubieten. Allerdings sollte den Studierenden die Frage nach der Selbsteinschätzung der Lesekompetenz in einer differenzierteren Form angeboten werden als bislang. Bei der globalen Einschätzung der Kompetenz ist es für Studierende schwierig nachzuvollziehen, welche Teildimensionen der Lesekompetenz bewertet werden sollen. Oft ist es der Fall, dass unter ‚Lesenkönnen‘ eine einfache Buchstaben-Laut-Zuordnung verstanden wird, die von den universitären Anforderungen an die Lesefähigkeiten der Studierenden deutlich abweicht.

Für zukünftige Untersuchungen könnten Bewertungsskalen für Studierende entwickelt werden. Ein gutes Beispiel der Bewertungsskala für sprachliche Fähigkeiten bietet GER und DIALANG (Selbstbewertungsskalen). Da diese Deskriptoren nur für den Fremdsprachenbereich gelten, könnten sie für die Gruppe der Studierenden angepasst werden.

Die Hypothese, dass Studierende, die ihre Lesekompetenz im Deutschen gut oder sehr gut einschätzen, im Durchschnitt höhere C-Testleistungen erzielen als Studierende, die ihre Lesekompetenz im Deutsch schlecht einschätzen, konnte nur teilweise bestätigt werden. In der Gesamtgruppe der Studierenden und in der Gruppe der Studierenden mit Migrationshintergrund besteht ein signifikanter mittlerer Zusammenhang zwischen der Selbsteinschätzung der Lesekompetenz und den C-Testleistungen. Allerdings lässt sich in der Gruppe der deutschen Muttersprachler kein signifikanter Zusammenhang nachweisen.

8.2.5 C-Testleistung und Selbsteinschätzung der schriftlichen Kompetenz

Hypothese 17: *Studierende, die ihre schriftliche Kompetenz im Deutschen gut oder sehr gut einschätzen, erzielen im Durchschnitt signifikant bessere C-Testleistungen als Studierende, die ihre schriftliche Kompetenz im Deutschen schlecht oder sehr schlecht einschätzen.*

Angenommen wird, dass sich produktive schriftsprachliche Kompetenzen durch die C-Tests miterfassen lassen und mit den selbst eingeschätzten Kompetenzen übereinstimmen.

Untersuchungen im fremdsprachlichen Bereich zur Überprüfung der Validität von C-Tests mit anderen Tests zu verschiedenen Teilkompetenzen (z. B. Schreiben und Lesen) bestätigten, dass C-Tests diese Bereiche miterfassen (vgl. Klein-Braley & Raatz, 1984; Klein-Braley, 1985; Sigott, 2004). Festgestellt wurde, dass fremdsprachliche C-Tests (z. B. in den untersuchten Sprachen Deutsch und Englisch) hohe Korrelationen mit produktiven schriftsprachlichen Teilkompetenzen aufweisen. Im Folgenden wird auf einige korrelative Studien eingegangen.

Eckes (2011b) stellte einen starken Zusammenhang zwischen den C-Tests onDaF und TestDaF im schriftlichen Ausdruck fest, die hohe Korrelationen nach Pearsons aufweisen ($r = 0.62$). C-Tests korrelieren auch relativ hoch mit den Bereichen Wortschatz und Grammatik, die ein fester Bestandteil jeder Schreibaufgabe sind, sowie mit den Ergebnissen aus Diktaten. Klein-Braley (1997) schrieb: "dictation is very often the best representative of the general factor" (S. 54) und "another indication that the C-Test measures in test batteries language proficiency" (Eckes & Grotjahn, 2006a, S. 299). Jafarpur (2002) beobachtete im English Placement Test (EPT) hohe Korrelationen zwischen der C-Testleistung und den Bereichen Grammatik ($r = 0.72$ bis 0.84) und Wortschatz ($r = 0.74$ bis 0.85) (vgl. S. 35f.). Zu einem ähnlichen Ergebnis kamen auch Babaii & Ansary (2001). Der Zusammenhang zwischen der Leistung im TOEFL-Test im Bereich Wortschatz und dem C-Testergebnis betrug $r = 0.79$. Daller et al. (2003)

berichteten, dass bei der erfolgreichen Rekonstruktion von C-Tests lexikalische Kompetenz eine wichtige Rolle spiele (vgl. Daller, 2003, S. 206).

Zwischen verschiedenen Subtests in der TestDaF-Prüfung wurden ebenso hohe Korrelationswerte festgestellt; die Korrelationen mit dem Bereich Schreiben betrugen $r = 0.68$ (vgl. Arras et al., 2002). Alle diese Studie belegen, dass C-Tests schriftliche Teilkompetenzen miterfassen.

Erklärt werden können die bestehenden Korrelationen dadurch, dass bei der Bearbeitung von C-Tests gleichzeitig mehrere sprachliche Teilkompetenzen eingesetzt werden müssen, da sie aufgrund ihrer Konstruktion, der das Prinzip der reduzierten Redundanz zugrunde liegt, (vgl. Kap. 3) als integrative Testverfahren verstanden werden können, die Sprache ganzheitlich erfassen. Das bedeutet, dass die sprachlichen Fähigkeiten nicht in vier Fertigkeiten (Hören, Sprechen, Lesen, Schreiben) oder andere spezifische Sprachbereiche, zum Beispiel Morphologie oder Semantik, zerlegt werden (dazu auch das Kap. 6.1), sondern global zu betrachten sind.

Bis jetzt liegen die Ergebnisse der Korrelationsstudien nur für fremdsprachliche C-Tests vor. Ein erster Versuch, schriftsprachliche Kompetenzen mithilfe von C-Tests zu erfassen, wurde im *Sprachkompetenzprojekt Diagnose und Förderung* an der Universität Duisburg-Essen vorgenommen. Zwischen den Leistungen aus den muttersprachlichen C-Tests und den durch die Schreibaufgabe gemessenen schriftlichen Kompetenzen der Studierenden¹⁴¹ wurde allerdings nur ein kleiner signifikanter Zusammenhang festgestellt (dazu Bremerich-Vos et al., 2013, auch Scholten-Akoun, Tischmeyer & Mashkovskaya, 2013). Angenommen wird, dass die für die Studierenden modifizierten C-Tests, die die Wortanfangstilgung enthalten, primär semantisch-lexikalischen Fertigkeiten erfassen. Aufgrund der oben genannten Punkte wird erwartet, dass die Selbsteinschätzung der schriftlichen Kompetenz eine Übereinstimmung mit den durch die C-Test eingeschätzten sprachlichen Leistungen haben wird, da Ergebnisse aus den C-Tests und schriftsprachliche Kompetenzen, erfasst durch die Schreibaufgaben, korrelieren.

¹⁴¹ Kombiniert wurden die C-Tests mit einer Schreibaufgabe, die von den Studierenden verlangte, einen argumentativen Text auf der Basis eines Zeitungsartikels zu verfassen.

8. Hypothesen: Ergebnisse und Kommentare

In der vorliegenden Studie konnten für die Berechnung der Korrelationen zwischen der C-Testleistung und der Selbsteinschätzung der schriftlichen Kompetenz die Daten von insgesamt 169 Studierenden analysiert werden, was ca. 19 % der getesteten Studierenden entspricht¹⁴². Zwischen beiden erhobenen Variablen wurde ein kleiner signifikanter Zusammenhang beobachtet; die berechnete Korrelation nach Pearson beträgt $r = -0.413$, $p = .000 < .01$ (Sig. 2-seitig). Der Korrelationskoeffizient wird als ein schwacher Zusammenhang interpretiert. Das bedeutet: Wird die schriftliche Kompetenz von den Studierenden als sehr gut eingeschätzt, heißt es nicht unbedingt, dass bei den C-Tests ein hoher Wert erreicht wird. Gleiches gilt für Studierende mit einer niedrigen Selbsteinschätzung der schriftlichen Kompetenz. Ein kleiner Zusammenhang bedeutet, dass es durchaus zu Überschneidungen zwischen den Einschätzungen auf der fünfstufigen Skala und den fünf Leistungsstufen, denen die C-Testergebnisse der Studierenden zugeordnet werden, kommen kann. Allerdings sind solche Überschneidungen nicht immer konstant.

Um zu überprüfen, ob sich die Ergebnisse von der Variable Migrationshintergrund beeinflussen lassen, werden zwei separate Gruppen gebildet: die Gruppe der Studierenden mit und ohne Migrationshintergrund. Auffallend ist, dass in der Gruppe der Studierenden mit der Muttersprache Deutsch ($N = 82$) etwas schwächere Korrelationen beobachtet werden als in der Gesamtgruppe. Der Korrelationskoeffizient nach Pearson beträgt $r = -0.299$ und ist signifikant, $p = 0.005 < .01$ (Sig. 2-seitig). Das bedeutet, dass zwischen den Variablen C-Testleistung und Selbsteinschätzung der schriftlichen Sprachkompetenz in der Gruppe der Studierenden ohne Migrationshintergrund ein schwacher positiver Zusammenhang besteht.

Berechnet man die Korrelationen für die Gruppe der Studierenden mit Migrationshintergrund ($N = 86$), so ist ein kleiner Anstieg der Korrelation zu beobachten. Der berechnete Korrelationskoeffizient nach Pearson beträgt $r = -0.358$ und ist auch signifikant, $p = 0.001 < .01$ (Sig. 2-seitig) (vgl. dazu

¹⁴² In der Erhebung konnten nicht die Daten aller getesteten Studierenden ($N=887$) genutzt werden, da die Frage nach der Einschätzung der schriftlichen Kompetenz erst im Endstadium der Erhebung in den Fragekatalog aufgenommen wurde und daher nicht allen getesteten Studierenden vorlag.

Berechnungen im Anhang A, Teil 3.2.5). Eine mögliche Erklärung für den Anstieg des Korrelationskoeffizienten in der Gruppe der Studierenden mit Migrationshintergrund wären unterschiedliche Varianzen in den getesteten Gruppen. In der Gruppe der Studierenden mit Migrationshintergrund streuen die C-Testleistungen innerhalb aller Leistungsstufen; in der Gruppe der Studierenden ohne Migrationshintergrund sind die Streuungsbereiche nicht so breit, was Auswirkungen auf die Korrelationen hat.

Zusammenfassend lässt sich festhalten, dass ein Zusammenhang zwischen der Selbsteinschätzung der schriftsprachlichen Kompetenzen und den C-Testergebnissen besteht – allerdings ist er etwas stärker in der Gruppe der Studierenden mit Migrationshintergrund, also derjenigen Studierenden, die eine andere Muttersprache und vermutlich Deutsch als Zweit- oder Fremdsprache erlernt haben. Die erhobenen Daten über die schriftliche Kompetenz liefern zusätzliche Informationen über die Selbsteinschätzung der Studierenden in Bezug auf ihre produktiven sprachlichen Leistungen, erscheinen aber aus zwei Gründen als problematisch: Zum einen kann es vorkommen, dass Studierende ihre schriftliche Kompetenz über- oder unterschätzen, und zum anderen werden die Selbsteinschätzungen von üblicherweise in der Schule bewerteten produktiven Leistungen (z. B. Klausuren, Hausarbeiten, Essays etc.) beeinflusst. Daher kann von einer hohen Selbsteinschätzung der schriftlichen Kompetenz durch Studierende nicht automatisch auf eine hohe erbrachte C-Testleistung geschlossen werden. Die schriftsprachlichen Teilkompetenzen werden durch die C-Tests miterfasst – allerdings nicht auf ähnliche Weise wie das z. B. in einem Essay stattfindet.

Die Hypothese, dass Studierende, die ihre schriftliche Kompetenz im Deutschen gut oder sehr gut einschätzen, im Durchschnitt signifikant bessere C-Testleistungen erzielen als Studierende, die ihre schriftliche Kompetenz im Deutschen schlecht oder sehr schlecht einschätzen, konnte bestätigt werden.

9. Abschließender Ausblick

Mit der vorliegenden Studie konnte gezeigt werden, dass mit dem Einsatz von adressatenspezifischen C-Tests im universitären Bereich nicht nur Aussagen über den Grad der allgemeinen Sprachkompetenz von Studierenden möglich sind, sondern auch Aussagen über die Lesefähigkeit.

Zur Entwicklung und Analyse von C-Tests wurden umfangreiche Vorarbeiten aus dem *Sprachkompetenzprojekt Diagnose & Förderung* an der Universität Duisburg-Essen verwendet, in dessen Rahmen die vorliegende Arbeit entstanden ist. Auf der Grundlage dieser Vorarbeiten konnte belegt werden, dass sich modifizierte C-Tests mit Wortanfangstilgung, die eine Grundlage für die vorliegende Arbeit bilden, am besten für Studierende eignen. Anders als C-Tests mit Wortendtilgung ermöglichen sie eine Verschiebung des Fokus' auf die Worterkennung und erfassen somit semantisch-lexikalische Fähigkeiten, was bei der Überprüfung des Leseverständnisses zentral ist (vgl. Kap. 4).

Können C-Tests die Lesefähigkeiten der Lehramtsstudierenden abbilden?

Wie die Ergebnisse der Studie belegen, können die entwickelten C-Tests sowohl zur Überprüfung der Leseprozesse auf den hierarchieniedrigen als auch auf den hierarchiehöheren Ebenen verwendet werden, die sich auch durch adressatenspezifische Lesetests abbilden lassen (vgl. Kap. 6). Den Beschreibungen des Leseprozesses auf der Lückenebene, die in der vorliegenden Studie durchgeführt wurden, liegt die Annahme zugrunde, dass bei der Rekonstruktion von C-Tests verschiedene Prozessebenen (Wort-, Satz- und Textebene) aktiviert werden (vgl. Kap. 2) – ähnlich wie beim Dekodieren und Verstehen von Lesetexten. Die Zuordnung der Lücken zu verschiedenen Prozessebenen liefert Aussagen über den Leseprozess, der sich bei der Vervollständigung der C-Testlücken vollzieht (vgl. Kap. 6). Die Durchführung der Lückenanalyse ermöglicht eine Differenzierung innerhalb der Gruppe, d. h. zwischen leistungsstarken und leistungsschwachen Studierenden. In der Studie wurde bestätigt, dass bei der Anwendung von Lesestrategien auf

unterschiedlichen Ebenen zwischen diesen Gruppen signifikante Unterschiede bestehen.

Überprüft und bestätigt in Bezug auf die Leseleistungen wurde die Hypothese, dass Studierende, die bereits bei der Rekonstruktion der Lücken auf Wortebene Schwierigkeiten haben, die Rekonstruktion der Lücken auf Satz- und Textebene größtenteils nicht gelingt. Dadurch wurde deutlich, dass für die erfolgreiche Rekonstruktion von C-Tests die Wortebene nicht ausreicht. Wenn die Studierenden die C-Tests auf höheren Prozessebenen lösen, rekonstruieren sie in der Regel auch Lücken auf niedrigeren Ebenen. Die Unterschiede zwischen leistungsstarken und leistungsschwachen Lesern steigen, sobald im C-Test Lücken auftauchen, die höheren Prozessebenen zugeordnet sind und Studierende textübergreifende Lesestrategien anwenden müssen. So sind die Leistungsunterschiede auf der Textebene am stärksten, da leistungsschwachen Lesern auf dieser Ebene vermutlich komplexe Lesestrategien fehlen, was zu großen Leistungsdifferenzen führt. Auf der Wortebene sind die Unterschiede am geringsten, da die Rekonstruktion der C-Test-Lücken auf dieser Ebene durch Heranziehung des näheren Kontextes und von Vorwissen erfolgt, was auch leistungsschwächere Studierende beherrschen.

Erhebung persönlicher Daten mit dem Fragebogen:

Durch den Einsatz des Fragebogens wurden zusätzliche Variablen erhoben, die persönliche Informationen z. B. in Bezug auf die Selbsteinschätzung der sprachlichen Kompetenz durch die Studierenden liefern (vgl. Kap. 8.1 und Kap. 8.2). Korrelationsstudien haben ergeben, dass die Selbsteinschätzung der Sprachkompetenz als aussagekräftiger zusätzlicher Prädiktor nicht nur für die C-Testleistungen, sondern auch für die Lesekompetenz verwendet werden kann. Zwischen der selbst eingeschätzten Lesekompetenz und den C-Tests wurden Korrelationen festgestellt.

Folgende weitere Hypothesen wurden im Rahmen der vorliegenden Studie *bestätigt*:

- C-Testleistungen männlicher und weiblicher Studierender unterscheiden sich signifikant voneinander.

- Studierende mit Migrationshintergrund erzielen im C-Test signifikant schlechtere Ergebnisse als Studierende ohne Migrationshintergrund.
- Studierende, die bereits eine Ausbildung oder ein Studium abgeschlossen haben, erzielen signifikant höhere C-Testergebnisse.
- Studierende mit einem Gymnasialabschluss erzielen signifikant höhere sprachliche Leistungen in den C-Tests als Absolventen anderer Schulformen.
- Studierende, die ihre Sprachkompetenz im Deutschen gut einschätzen, erzielen signifikant bessere C-Testleistungen.
- Studierende, die ihre schriftliche Kompetenz im Deutschen gut einschätzen, erzielen signifikant bessere C-Testleistungen als Studierende, die ihre Kompetenz schlechter einschätzen.
- Studierende, die ihre Lesekompetenz im Deutschen gut einschätzen, schätzen auch ihre schriftliche Kompetenz gut ein.

Überprüfung der Korrelationen zwischen den C-Tests und Lesetests:

Zusammen mit den C-Tests wurden verschiedene Lesetests eingesetzt – zum einen, um zu überprüfen, wie hoch der korrelative Zusammenhang zwischen C-Tests und den verwendeten Lesetests ist und zum anderen, um zu untersuchen, welche Determinanten der Lesekompetenz C-Tests und Lesetests gemeinsam abbilden können. Durch statistische Analysen wurde bestätigt, dass zwischen C-Tests und Lesetests je nach Lesetestformat mittlere bis hohe Korrelationen vorliegen (vgl. Kap. 6), was darauf schließen lässt, dass C-Tests zur Überprüfung der Lesefähigkeit von Studierenden als *geeignet* eingeschätzt werden können. Möchte man Aussagen über die Lesefähigkeit und die Anwendung von Lesestrategien von Studierenden machen, eignet sich dafür die Prozessebenenanalyse, die über die Kategorisierung der C-Test-Lücken erreicht werden kann (vgl. Kap. 6). Werden die C-Test-Lücken vor der Testung verschiedenen Kategorien zugeordnet (Wort-, Satz- und Textebene), können Voreinschätzungen bezüglich der Lesefähigkeiten und Lesestrategien von Studierenden gemacht werden. Ordnet man die C-Testleistungen verschiedenen Leistungsstufen zu (vgl. Kap. 7), können Leseleistungen, die niedrigeren Stufen

zugeordnet sind, zusätzlich auf die Beherrschung von Lesestrategien auf der Wort-, Satz- und Textebene hin untersucht werden.

Scheitern Studierende bei der Bearbeitung auf der Textebene bei der Rekonstruktion von C-Tests, können weitere Leseverständnistests, z. B. Stolperwörtertests für Erwachsene (vgl. Kap. 5), genutzt werden, um zu überprüfen, ob die Betreffenden Prozesse auf den hierarchieniedrigeren Ebenen beherrschen.

Die Kategorisierung der Lücken bietet die Möglichkeit, den C-Test nicht nur zur Überprüfung der allgemeinen Sprachkompetenz einzusetzen und mit einem *Globalwert* zu operieren, sondern auch Teilkompetenzen wie das Lesen zu messen und die erzielten Werte in Bezug auf die Leseleistungen hin zu interpretieren. Der Einsatz von C-Tests mit der Wortanfangstilgung und die Analyse von C-Testlücken mit dem Fokus auf lexikalisch-semantischen Kenntnissen können Anhaltspunkte über spezifische Teilkompetenzen im Bereich Lesen liefern, auf deren Grundlage diagnostisch verwertbare Aussagen gemacht werden können.

Die beschriebenen Befunde stellen für die C-Test-Forschung eine Erweiterung dar, da C-Tests mit Lesetests bis jetzt nur im fremdsprachlichen Bereich korreliert wurden und Lückenanalysen vor allem auf die Bestimmung der Schwierigkeit von C-Test-Texten ausgerichtet waren (vgl. Kap. 4). Die vorliegenden Ergebnisse entkräften auch die Kritik, dass C-Tests kein Leseverständnis messen können (vgl. Kap. 4).

Konsequenzen für das Sprachkompetenzprojekt Diagnose & Förderung:

Im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* an der Universität Duisburg-Essen könnten die beschriebenen Befunde der vorliegenden Studie bedeuten, dass eine Abkopplung der C-Tests von der Schreibaufgabe möglich ist. Die Ergebnisse zeigen, dass C-Tests zur Überprüfung verschiedener Teildimensionen der Lesekompetenz eingesetzt werden können (vgl. Kap. 2, Kap. 6).

Dem Einsatz von C-Tests und der Schreibaufgabe im Sprachkompetenzprojekt liegt die Annahme zugrunde, dass durch C-Tests schriftsprachliche Fähigkeiten

gemessen werden können. Diese Annahme beruht auf zahlreichen korrelativen Untersuchungen zum C-Test im fremdsprachlichen Bereich, die belegen, dass der C-Test Teildimensionen der Schreibkompetenz wie zum Beispiel Orthographie, Grammatik, Syntax oder Kohärenz miterfasst.

Im Sprachkompetenzprojekt wurden Korrelationen berechnet (vgl. Publikation von Bremerich-Vos et al., 2013), um zu überprüfen, ob der C-Test zur Voreinschätzung der schriftsprachlichen Fähigkeiten der Studierenden in der universitären Praxis als valider Prädiktor genutzt werden kann und ob diese Voreinschätzungen mit den Einschätzungen durch die Schreibaufgabe übereinstimmen.

Niedrige Korrelationen zwischen beiden Testinstrumenten, die im Sprachkompetenzprojekt berechnet wurden, zeigten, dass durch die Verwendung von C-Tests bei Studierenden andere Teilkompetenzen fokussiert werden als die, die durch die Schreibaufgabe abgebildet werden. Diese These müsste allerdings noch genauer überprüft werden. Zwischen beiden Testinstrumenten wurden jedenfalls nur schwache Korrelationen beobachtet, was folgende Gründe haben könnte:

Zum einen können schwache Korrelationen dadurch entstehen, dass C-Tests und die Schreibaufgabe *verschiedene* Konstrukte darstellen. Die C-Tests erfassen stärker rezeptive Fähigkeiten, die Schreibaufgabe stärker produktive Fähigkeiten der Studierenden.

Zum anderen kann es durchaus sein, dass die C-Tests und die Schreibaufgabe *unterschiedliche Dimensionen* der Sprachkompetenz messen und zudem auch noch auf *unterschiedliche Weise*.

In den C-Tests wird den Studierenden ein enger Kontext zur Verfügung gestellt, und sie rekonstruieren nur das, was im Rahmen der Wortelisionen vorgegeben ist. Im Rahmen der Schreibaufgabe ist ebenfalls ein Kontext gegeben, wird allerdings nicht so eng gefasst. Den Studierenden wird ein Schreibimpuls in Form einer Aufgabe gegeben (z. B. Verfassen eines argumentativen Textes auf Grundlage eines Zeitungsartikels). Bestimmte Anforderungen steuern die Gestaltung dieses Textes – zum Beispiel ist der Leser des zu schreibenden Textes der Dozent.

Beide Testinstrumente haben vermutlich einen unterschiedlichen Schwierigkeitsgrad und erfordern die Aktivierung unterschiedlicher Strategien. Es könnte durchaus sein, dass sich bei Wahl einer anderen Textsorte als Schreibimpuls andere Korrelationen ergeben – zum Beispiel, wenn der Text keine starke Argumentationsstruktur aufweist, die wiedergegeben werden soll.

Da die Schreibaufgabe eine kombinierte Testaufgabe darstellt, in der nicht nur die Schreibfähigkeit fokussiert wird, sondern auch die Lesefähigkeit (Text- und Leseverständnis), können bei der Bewertung der Schreibaufgabe Konfundierungsprobleme auftauchen. Es ist möglich, dass Studierende den Ausgangstext gut verstanden haben, im geschriebenen Text aber aufgrund fehlender argumentativer Struktur und grammatischer oder orthografischer Fehler schwächere Leistungen erzielen. Es ist aber auch möglich, dass den Studierenden aufgrund fehlenden Leseverständnisses die Produktion eines Textes, der universitären Anforderungen entspricht, nicht gelingt. In beiden Fällen erzielen Studierende wenige Punkte – doch dafür liegen unterschiedliche Gründe vor.

Denkbar wäre der Einsatz einer anderen Schreibaufgabe, die zusammen mit den C-Tests eingesetzt wird. So könnten andere als die bisher verwendeten Textquellen und Themen genutzt werden. Eine mögliche Alternative wäre eine Auswahl der Texte aus anspruchsvollen Fachzeitschriften wie *Psychologie heute* oder *Pädagogik*, die Themen beleuchten, die in Bezug auf Studium, Beruf (nicht nur Lehrerberuf) oder zwischenmenschliche Beziehungen für Studierende interessant sind. Auf Grundlage dieser Texte könnten Studierende nicht nur eine inhaltliche Zusammenfassung des gelesenen Artikels vornehmen, sondern auch Stellung zu einem aktuellen Problem beziehen. Der Schreibimpuls könnte auch Platz für persönliche Erfahrungen zum gelesenen Thema bieten, was den motivationalen Aspekt der Schreibaufgabe erhöhen würde. Da die Motivation der Studierenden eine wichtige Rolle in Bezug auf die Validität der Ermittlung und Interpretation der Ergebnisse und auf die Testung insgesamt spielt, sollte dieser Aspekt vor dem Einsatz der Schreibaufgabe gut bedacht werden. Durch Verwendung einer modifizierten Schreibaufgabe könnten nicht zuletzt auch andere Korrelationen entstehen, da sowohl die Aufgabenstellung als auch die

Textsorte oder das Thema die Schwierigkeit der Schreibaufgabe für die Studierenden in gewisser Weise mitbestimmen.

Konsequenzen für die Sprachdiagnostik:

Mit der vorliegenden Studie konnte gezeigt werden, dass die Anwendung des C-Testformats nicht nur in Bezug auf die neue Zielgruppe (*Lehramtsstudierende im Grundstudium*), sondern auch auf den neuen Testbereich bei Muttersprachlern (*Lesekompetenz*) ausgeweitet werden kann. Außerdem bieten C-Tests die Möglichkeit, gemischte Gruppen von Studierenden zu testen, das heißt, mit den entwickelten C-Tests lassen sich Sprachfähigkeiten von Studierenden mit und ohne Migrationshintergrund erfassen. Für die vorliegende Zielgruppe wurden Normwerte ermittelt, an denen die Leistungen der Lehramtsstudierenden an der Universität Duisburg-Essen gemessen werden können (vgl. Kap. 7). Die ermittelten Normwerte sind stabil und repräsentativ, da den Berechnungen die Daten von ca. 2.000 Lehramtsstudierenden zugrunde liegen (vgl. dazu Publikation von Bremerich-Vos et al., 2013).

Für sprachdiagnostische Zwecke wären in Bezug auf die Testung von Studierenden folgende Befunde der Studie *besonders wichtig*:

Der C-Test hat sich auch in dieser Studie als ein sehr ökonomisches und praktikables Testinstrument erwiesen, dessen Einsatz im universitären Kontext aus mehreren Gründen empfehlenswert ist. Zum einen ist es im Rahmen des *Sprachkompetenzprojektes Diagnose & Förderung* und der vorliegenden Arbeit gelungen, adressatenspezifische C-Tests zu entwickeln, die für die Zielgruppe mittelschwer sind, sodass eine Differenzierung zwischen leistungsstarken und leistungsschwachen Studierenden möglich wird. Zum anderen wurden zwischen den C-Tests und den Leseverständnistests mittlere bis hohe Korrelationen bestätigt (vgl. Kap. 6.4.1 und Kap. 6.4.2), was darauf schließen lässt, dass sich mit dem C-Test auf ökonomische Weise Lesefähigkeiten der Studierenden erfassen lassen. Durch eine objektive und ökonomische Auswertung kann den Studierenden eine schnelle Rückmeldung gegeben werden, was im universitären Kontext für Studierende und Dozenten sehr wichtig ist.

Ökonomische Überlegungen in Bezug auf Durchführung und Auswertung von C-Tests lassen es sinnvoll erscheinen, das Testformat nicht nur als 'Paper and Pencil'-Testung durchzuführen, sondern es auch als Online-Variante anzubieten, was im Rahmen universitärer Online-Prüfungen attraktiv ist. Durch Vorprogrammierung der C-Test-Lösungen lässt sich eine automatische Auswertung von C-Tests durchführen, was bei der Schreibaufgabe zum Beispiel nicht möglich ist.

Eine onlinebasierte Durchführung bietet vor allem im Rahmen der universitären Ausbildung die Möglichkeit, sehr große Gruppen von Studierenden zu testen, was beim Einsatz von C-Tests zum einen für Repräsentativität der Daten sorgt und die Datengrundlage erweitert und zum anderen zur Stabilisierung der ermittelten Normwerte beiträgt. Diese könnten dadurch an Generalisierbarkeit der Interpretation der Ergebnisse gewinnen.

Mit der vorliegenden Studie konnte insgesamt gezeigt werden, dass der C-Test als Format auf unterschiedliche Zielgruppen und für verschiedene Testzwecke übertragbar ist und in dieser Arbeit als valider Indikator zur Überprüfung der Lesekompetenz von Studierenden gelten kann.

Weiterer Forschungsbedarf:

Da die Studie Korrelationen zwischen den C-Tests und nur bestimmten Sorten von Leseverständnistests bestätigt hat, könnten in weiteren Untersuchungen andere Lesetestaufgaben zusammen mit den C-Tests getestet werden. So könnten z. B. adressatenspezifische MC-Aufgaben entwickelt werden, die dem sprachlichen Niveau und Wissensstand der Studierenden entsprechen. Um die Entwicklung der Lesefähigkeiten von Studierenden zu beobachten (besonders von Studierenden, die schwache Leistungen in den C-Tests und Lesetests erzielen), wäre die Messung der Sprachkompetenz der Lehramtsstudierenden im Grund- und Hauptstudium notwendig. Dies ließe sich in einer Längsschnittstudie realisieren, wenn man gleiche Studierende zweimal testen würde, z. B. bei Anfang und Ende ihres Studiums.

Besondere Aufmerksamkeit erfordern Studierende mit Migrationshintergrund, die, wie diese Studie und viele andere Studien (vgl. Kap. 8) belegen, signifikant

schwächere C-Testleistungen produzieren und im Studium häufiger Schwierigkeiten beim Verfassen schriftlicher Arbeiten haben. Das bedeutet aber nicht, dass diese Schwierigkeiten nur im Bereich *Schreiben* vorliegen. *Lesefähigkeiten* werden im Studium nicht gemessen, da davon ausgegangen wird, dass sie mit Aufnahme des Studiums an einer Universität bereits weit entwickelt sind.

Schriftliche Fähigkeiten entwickeln sich im Einklang mit rezeptiven Fähigkeiten im Bereich Lesen und bilden die Grundlage für akademische Fähigkeiten und die Studierfähigkeit insgesamt. Ausgeprägte Kompetenzen sowohl in der sprachlichen Teilfertigkeit Lesen als auch in der des Schreibens müssen die Studierenden für ihr Studium ausbilden (oder noch besser: in der Schule bereits ausgebildet haben). Es erscheint wichtig, an dieser Stelle zu betonen, dass Schwierigkeiten, die beim Leseprozess bei Studierenden vorhanden sind, früh genug diagnostiziert werden sollten, um adressatenspezifisch gefördert werden zu können. Für die Diagnose eignen sich die in dieser Studie vorgestellten C-Tests mit der Wortanfangstilgung.

Die beschriebenen Teilbereiche Lesen und Schreiben sind, wie bereits beschrieben, nicht als unabhängig voneinander zu betrachten, sondern bilden im Sinne der sprachlichen Mehrdimensionalität eine Basis für die übergeordnete Sprachkompetenz (vgl. Kap. 1). Es bleibt weiterhin zu untersuchen, ob diese Teilbereiche der Sprachkompetenz sich durch den Einsatz von C-Tests ähnlich beschreiben lassen. Besonderes Augenmerk wäre weiterhin auf den schriftsprachlichen Bereich zu richten, der in Verbindung mit dem C-Test für Studierende noch nicht ausreichend untersucht wurde. Ließe sich ein Zusammenhang zwischen den C-Tests und der Schreibaufgabe nachweisen, so hätte man einen validen Prädiktor in Form von C-Tests, mit dem sich die rezeptiven und produktiven Fähigkeiten der Studierenden auf ökonomische Weise einschätzen lassen. Der C-Test würde natürlich nicht die Schreibproduktion in Form von Essays und Aufsätzen ersetzen, böte aber ein praktikables Instrument, das im universitären Kontext breite Verwendung finden könnte. Möchte man die Schreibaufgabe mit den C-Tests zusammen testen, müsste eine Schreibaufgabe gefunden werden, die eine ökonomische und für die

universitäre Anwendung praktikable Auswertung ermöglicht. Die im *Sprachkompetenzprojekt* angebotenen Schreibimpulse wären für die Adressatengruppe der Studierenden in Bezug auf die Auswertung sehr komplex. Hier wäre, wie bereits ausgeführt, eine andere Auswahl der Grundlagentexte notwendig.

Hervorzuheben im Rahmen dieser Diskussion wären noch die in den VERA- und DESI-Studien beschriebenen Testkonstrukte (vgl. Kap. 1 und Kap. 3), deren Anregungen für veränderte Testkonstruktionen im Rahmen weiterer Vergleichsforschungen mit dem C-Test genutzt werden könnten.

Literaturverzeichnis

Adams, B.; Bell, L. C. & Perfetti, C. A. (1995): A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse Processes*, 20, p. 307-323.

Alderson, J. C. (1978): A study of the cloze procedure with native and non-native speakers of English. Unpublished PhD Dissertation, University of Edinburgh.

Alderson, J. C. & Urquhart, A. H. (Hrsg.) (1984): *Reading in a foreign language*. London: Longmann.

Arras, U. & Grotjahn, R. (2002): TestDaF: Aktuelle Entwicklungen. *Fremdsprachen und Hochschule*, 66, S. 65–88.

Arras, U.; Eckes, T. & Grotjahn, R. (2002): C-Tests im Rahmen des „Test Deutsch als Fremdsprache“ (TestDaF): Erste Forschungsergebnisse. In: R. Grotjahn (Hrsg.): *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*, Bd. 4, S. 175–209. Bochum: AKS-Verlag.

Arras, U. (2006): Testen und Beurteilen des Leseverstehens in der Fremdsprache. *Babylonia*, 6 (3–4), S. 81–86.

Arras, U. (2007): Wie beurteilen wir Leistung in der Fremdsprache? Strategien und Prozesse bei der Beurteilung schriftlicher Leistungen in der Fremdsprache am Beispiel der Prüfung Test Deutsch als Fremdsprache (TestDaF). Tübingen: Narr.

Artelt, C. (2000): *Strategisches Lernen*. Münster: Waxmann.

Artelt, C.; McElvany, N.; Christmann, U.; Richter, T.; Groeben, N.; Köster, J.; Schneider, W.; Stanat, P.; Ostermeier, C.; Schiefele, U.; Valtin, R. & Ring, K. (2007): *Expertise. Förderung der Lesekompetenz*, Bildungsforschung, Bd. 17. Bonn, Berlin, http://www.bmbf.de/pub/bildungsreform_band_siebzehn.pdf.

Artelt, C.; Schiefele, U. & Schneider, W. (2001): Predictors of reading literacy. *European Journal of Psychology of Education*, 16, p. 363-383.

Association of Language Testers in Europe (ALTE) (2005): *Handreichungen für Testautoren*. Deutsche Übersetzung von Dr. S. Bolton, August 2006, http://www.testdaf.de/aktuelle/pdf/ALTE_Deutsche_HR_Vorwort.pdf.

Babaii, E. & Ansary, H. (2001): The C-test: a valid operationalization of reduced redundancy principle? *System*, 29, p. 209-219.

Bachman, L. F. & Palmer A. S. (1982): The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16 (4), p. 449-465.

Bachman, L. F. & Palmer, A. S. (1996): Language testing in practice: Designing and developing usefull language tests. Oxford: Oxford University Press.

Bachman, L. F. & Palmer, A. S. (1981): The construct validation of the FS1 oral interview. *Language Learning*, 31(1), p. 67-86.

Bachman, L. F. & Palmer, A. S. (1982): The construct of some components of communicative proficiency. *TESOL Quarterly*, 16 (4), p. 440-465.

Bachman, L. F. (1990a): *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. F. (1990b): Constructing measures and measuring constructs. In: Harley, B.; Allen; Cummins, J. & Swain, N. (Eds.): *The development of second language proficiency*, p. 26-38. Cambridge: Cambridge University Press.

Backhaus, A.; Brügelmann, H.; Knorre, S. & Metze, W. (2003): *Stolperwörter-Lesetest (Stolper- 1)*. Forschungsmanual. Testmanual, <http://www2.agprim.uni-siegen.de/lust/stolpermanual.pdf>.

Baitsch, C.; Jutzi, K.; Delbrouck, I. & Hasenenbein, U. (1998): Organisationales Lernen: Eine organisationspsychologische Konzipierung der Entwicklung von Kompetenz bei Individuen, Gruppen und Organisationen. In: H. Geißler; A. Lehnhoff; P. Petersen (Hrsg.): *Organisationslernen im interdisziplinären Dialog*, S. 91-108. Weinheim: Deutscher Studien Verlag.

Baur, R. S. & Meder, G. (1994): C-Tests zur Ermittlung der globalen Sprachfähigkeit im Deutschen und einer Muttersprache bei ausländischen Schülern in der Bundesrepublik Deutschland. In: R. Grotjahn (Hrsg.): *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Manuskripte zur Sprachlehrforschung, Band 39 (2), S. 151-178. Bochum: Brockmeyer.

Baur, R. S.; Chlosta, C. & Spettmann, M. (2005): Bericht der Arbeitsgruppe 4: Lehren, Lernen, Beurteilen. In: D. Schäffer (Hrsg.): *Deutsch als Fremdsprache in Europa. 2. Jahresbericht des Comenius-Netzwerks "DaF-Südost" 2004-2005*, S. 91-108. Athen, <http://www.daf-netzwerk.org/infos/jahresberichte/DaF-annual-report-2004-2005.pdf>.

Baur, R. S. & Spettmann, M. (2005): Tests im Rahmen des Gemeinsamen Europäischen Referenzrahmens für DaF. <http://www.daf-netzwerk.org/tagungen/ergebnisse/workshop2005-03/Tests-GER-DaF.doc>.

Baur, R. S. & Spettmann, M. (2006): Kompetenzstufen testen – leicht gemacht: C-Tests für DaF in der Praxis. In: D. Schäffer & M. Adamopoulou (Hrsg.): *Sprachen – Kulturen – Identität: Schule und Fortbildung für Europäer von morgen*, S. 149-161. Pallini, Griechenland: Ellinogermaniki Agogi.

- Baur, R. S. & Spettmann, M. (2007): Kompetenzen testen – leicht gemacht. C-Tests für die Orientierungsstufe. In: C. Bainski & M. Krüger-Potratz (Hrsg.): Handbuch Sprachförderung, S. 123-131. Essen: NDS.
- Baur, R. S. & Spettmann, M. (2007): Screening – Diagnose – Förderung: Der C-Test im Bereich DaZ. In: B. Ahrenholz (Hrsg.): Deutsch als Zweitsprache. Voraussetzungen und Konzepte für die Förderung von Kindern und Jugendlichen mit Migrationshintergrund, S. 95-110. Freiburg i.Br.: Fillibach.
- Baur, R. S. & Spettmann, M. (2008): Sprachstandsmessung und Sprachförderung mit dem C-Test. In: B. Ahrenholz & I. Oomen-Welke (Hrsg.): Deutsch als Zweitsprache (Deutschunterricht in Theorie und Praxis, hg. v. Winfried Ulrich, Band VIII), S. 430-441. Baltmannsweiler: Schneider-Hohengehren.
- Baur, R. S. & Spettmann, M. (2009): Der C-Test als Instrument der Sprachdiagnose und Sprachförderung. In: H.-J. Roth; Hans H. Reich & D. Lengyel (Hrsg.): Von der Sprachdiagnose zur Sprachförderung. FörMig Edition Bd. 5, S. 115-127. Münster: Waxmann.
- Baur, R. S. & Spettmann, M. (2010): Lesefertigkeiten testen und fördern. In: C. Benholz; G. Kniffka & E. Winters-Ohle (Hrsg.): Bildungssprachliche Kompetenzen fördern in der Zweitsprache. Beiträge des Mercator-Symposiums „Fachliche und sprachliche Förderung von Schülern mit Migrationsgeschichte“ im Rahmen des AILA-Weltkongresses 2008 in Essen, S. 95-114. Münster: Waxmann.
- Baur, R. S.; Grotjahn, R. & Spettmann, M. (2006): Der C-Test als Instrument der Sprachstandserhebung und Sprachförderung. In: J.-P. Timm (Hrsg.): Fremdsprachenlernen und Fremdsprachenforschung: Kompetenzen, Standards, Lernformen, Evaluation, S. 389-406. Tübingen: Narr.
- Baur, R. S. & Spettmann, M. (2006a): Kompetenzstufen testen - leicht gemacht. C-Tests für DaF in der Praxis. In: D. Schäffer & M. Adamopoulou (Hrsg.): Sprachen - Kulturen - Identität: Schule und Fortbildung für Europäer von morgen, S. 149-161. Pallini, Griechenland: Ellinogermaniki Agogi.
- Baur, R. S.; Grotjahn, R. & Spettmann, M. (2006b): Der C-Test als Instrument der Sprachstandserhebung und Sprachförderung. In: T. Johannes-Peter (Hrsg.): Fremdsprachenlernen und Fremdsprachenforschung: Kompetenzen, Standards, Lernformen, Evaluation, S. 389-406. Tübingen: Narr.
- Baur, R. S.; Mashkovskaya, A. & Weirich, S. (2013, in Vorbereitung): Ergebnisse und Diskussion der C-Tests. In: A. Bremerich-Vos; R. S. Baur & D. Scholten-Akoun (2013, in Vorbereitung): Erfassung der schriftsprachlichen Kompetenzen von Lehramtsstudierenden. Befunde aus dem Sprachkompetenzprojekt Diagnose & Förderung an der Universität Duisburg-Essen.

Baur, R. S.; Mashkovskaya, A. & Spettmann, M. (2010): Der C-Test als Instrument zur Ermittlung allgemeinsprachlicher und fachsprachlicher Fähigkeiten am Berufskolleg. In: A. Berndt & K. Kleppin (Hrsg.): Sprachlehrforschung: Theorie und Empirie, S. 23-38. Festschrift für Rüdiger Grotjahn. Frankfurt a. M.: Peter Lang.

Baur, R. S. & Scholten-Akoun, D. in Zusammenarbeit mit der Stiftung Mercator (Hrsg.) (2010): Deutsch als Zweitsprache in der Lehrerbildung: Bedarf – Umsetzung – Perspektiven,
www.mercator-foerderungunterricht.de/projekt/publikationen.html.

Baur, R. S. & Mashkovskaya, A. (2013, in Vorbereitung): C-Test-Kritik: revised (Arbeitstitel). In: Festschrift für K. Kleppin.

Beck, B. & Klieme, E. (Hrsg.) (2007): Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie (Deutsch Englisch Schülerleistungen International). Weinheim: Beltz.

Becker-Mrotzek, M. (1995): Wie entwickelt sich die Schreibfertigkeit? In: Diskussion Deutsch 141, S. 25-35.

Becker-Mrotzek, M. (1997): Warum Schreiben schwer ist. In: Becker-Mrotzek, M.; Hein, J. & Koch, H. H. (Hrsg.): Werkstattbuch Deutsch: Texte für das Studium des Faches, Bd. 1, S. 95-104. Münster: Lit-Verlag.

Becker-Mrotzek, M. (2003): Texte produzieren. In: G. Czech & J. Volmert (Hrsg.): Schreiben im integrativen Unterricht, S. 52-60. Baltmannsweiler: Schneider.

Bisping, M. & Raatz, U. (2002): Sind computerisierte und Papier & Bleistift-Versionen des C-Tests äquivalent? In: R. Grotjahn (Hrsg.): Der C-Test: theoretische Grundlagen und praktische Anwendungen, Bd. 4, S. 131-155. Bochum: AKS-Verlag.

Bundesministerium für Bildung und Forschung (BMBF) (2004): Bildung auf einen Blick - Wesentliche Aussagen (PDF),
http://www.bmbf.de/pub/bildung_auf_einen_blick_wesentliche_aussagen.pdf.

Bolton, S. (1996): Probleme der Leistungsmessung: Lernfortschrittstests in der Grundstufe. Studieneinheit 10 des Fernstudienprojekts „Deutsch als Fremdsprache und Germanistik“, Berlin: Langenscheidt.

Borgartz, S. & Raatz, U. (2002): „Sind C-Tests trainierbar?“ In: R. Grotjahn (Hrsg.): Der C-Test: theoretische Grundlagen und praktische Anwendungen, Bd. 4, S. 157-174. Bochum: AKS-Verlag.

Borkowski, J. G.; Schneider, W. & Pressley, M. (1989): The challenges of teaching good information processing to learning disabled students. International Journal of Disability, Development and Education, 36, p. 169-185.

Börner, W. & Vogel, K. (Hrsg.) (2004): Emotion und Kognition im Fremdsprachenunterricht. Tübingen: Gunter Narr Verlag.

Bortz, J. & Döring, N. (2006): Forschungsmethoden und Evaluation: für Human und Sozialwissenschaftler. Springer-Lehrbuch. Berlin, Heidelberg: Springer Medizin Verlag Heidelberg.

Bos, W.; Strietholt, R.; Goy, M.; Stubbe, T. C.; Tarelli, I. & Hornberg, S. (2011): IGLU 2006. Dokumentation der Erhebungsinstrumente. Münster: Waxmann.

Bremerich-Vos, A.; Baur, R. S. & Scholten-Akoun, D. (2013, in Vorbereitung): Erfassung der schriftsprachlichen Kompetenzen von Lehramtsstudierenden. Befunde aus dem Sprachkompetenzprojekt Diagnose & Förderung an der Universität Duisburg-Essen.

Bremerich-Vos, A.; Groß Ophoff, J.; Hosenfeld, I., Helmke & A. Wagner, S. (2005): Stellungnahme zur Kritik an VERA in „Grundschule aktuell“, 89. In: H. Bartnitzky (Hrsg.): Grundschule aktuell, 90, S. 3-6. Grundschulverband - Arbeitskreis Grundschule e.V., ergänzend, „Stellungnahme zur Kritik an VERA in »Grundschule aktuell«, Heft 89“ (Langversion).

Brügelmann, H. (2004): Leseleistungen von HandwerkerInnen im Stolperwörter-Lesetest. Erste Befunde und ihre Deutung. Projekt LUST/FB 2 der Universität: Siegen, www.uni-siegen.de/~agprim/lust.

Bühner, M. (2004): Einführung in die Test- und Fragebogenkonstruktion. München [u.a.]: Pearson Studium.

Canale, M & Swain, M. (1981): A theoretical framework for communicative competence. In: A. S. Palmer; P. G. Groot & S. A. Trosper (Eds.): The construct validation of tests of communicative competence, p. 31-26. Washington, D. C.: TESOL.

Carle, U. (2003): Kernkompetenzen von Lehrer/innen - empirische Befunde als Basis für Lehrerbildungsstandards? In: Landesinstitut für Lehrerbildung Bremen (Hrsg): Reform der Lehrerbildung. Dokumentation der zweiten Expertentagung Lehrerbildung vom 7. - 9.11.02, S.27-56. Bremen: LIS.

Carroll, J. B., Carton, A. S. & Wilds, C. P. (1959): An investigation of cloze items in the measurement of achievement in foreign languages. College Entrance Examination Board research and Development Report. Cambridge, Mass.: Laboratory for Research in Instruction, Harvard University.

Carroll, J.B. (1961): Fundamental considerations in testing for English language proficiency of foreign students. In: Testing proficiency of foreign students. Washington, D. C.: Center for Applied Linguistics. Also reprinted in H. B. Allen & R.N. Campbell (Eds.): Teaching English as a second language: A book of reading, 2nd ed., 1972. New York: McGraw-Hill Book Company.

Chlosta, C. & Ostermann, T. (2005): "Warum fragt man nach Herkunft, wenn man die Sprache meint? Ein Plädoyer für eine Aufnahme sprachbezogener Fragen in demographische Untersuchungen." In: Bundesministerium für Bildung und Forschung (BMBF) (Hrsg.): Migrationshintergrund von Kindern und Jugendlichen: Wege zur Weiterentwicklung der amtlichen Statistik, S. 55-66.

Chomsky, N. (1965): *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. (2001): *The architecture of language*. New Delhi: Oxford University Press.

Christmann, U. & Groeben, N. (1996): Textverstehen, Textverständlichkeit – Ein Forschungsüberblick unter Anwendungsperspektive. In: H. P. Krings (Hrsg.): *Wissenschaftliche Grundlagen der Technischen Kommunikation*, S. 129-189. Tübingen: Narr.

Christmann, U. & Groeben, N. (1999): Psychologie des Lesens. In: B. Franzmann, K. Hasemann; D. Löffler; E. Schön (Hrsg.): *Handbuch Lesen*, S. 145-223. München: Saur.

Cummins, J. (1979): Cognitiv/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Paper in Bilingualism*, 19, p. 97-205.

Daller, H.; van Hout, R. & Treffers-Daller, J. (2003): Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24, p. 197-222.

Daller, H.; Phelan, D. & Grotjahn, R. (2006): The C-test and TOEIC® as measures of students' progress in intensive short courses in EFL. In: R. Grotjahn (Ed.): *Der C-Test: Theorie, Empirie, Anwendungen/The C-test: theory, empirical research, applications*, S. 101-119. Frankfurt am Main: Peter Lang.

Darnell, D. K. (1968): *The development of an English Language proficiency test of foreign students using a clozentropy procedure*. Boulder, Co.: Department of Speech and Drama, University of Colorado.

Deci, E. L. & Ryan, R. M. (1985): *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.

DESI-Konsortium (2008): *Unterricht und Kompetenzerwerb. Zentrale Ergebnisse der Studie Deutsch-Englisch-Schülerleistungen International*. Frankfurt a. M.: DIPF.

Deutsches PISA-Konsortium (Hrsg.) (2000): *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

Dörnyei, Z. & Schmidt, R. (Hrsg.) (2001): *Motivation and second language acquisition*. Honolulu, HI: University of Hawaii Press.

Dresselhaus, G. (1979): *Langue, Parole und Kompetenz, Performanz: zur Klärung der Begriffspaare bei Saussure und Chomsky; ihre Vorgeschichte und ihre Bedeutung für die moderne Linguistik*, Frankfurt a. M. [u.a.]: Lang.

Eckes, T. (2003): *Qualitätssicherung beim TestDaF: Konzepte, Methoden, Ergebnisse*. Fremdsprachen und Hochschule, 69, S. 43–68.

Eckes, T. (2006): Rasch-Modelle zur C-Test-Skalierung. In: R. Grotjahn (Hrsg.): *Der C-Test: Theorie, Empirie, Anwendungen/The C-test: Theory, empirical research, applications*, S. 1–44. Frankfurt: Lang.

Eckes, T. & Grotjahn, R. (2006a): A closer look at the construct validity of C-tests. *Language Testing*, 23, p. 290–325.

Eckes, T. & Grotjahn, R. (2006b): C-Tests als Anker für TestDaF: Rasch-Analysen mit dem kontinuierlichen Ratingskalen-Modell. In: R. Grotjahn (Hrsg.): *Der C-Test: Theorie, Empirie, Anwendungen/The C-test: Theory, empirical research, applications*, S. 167–193. Frankfurt: Lang.

Eckes, T. (2007): Konstruktion und Analyse von C-Tests mit Ratingskalen-Rasch-Modellen. *Diagnostica*, 53, 68–82.

Eckes, T. (2010a): Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In: R. Grotjahn (Hrsg.): *Der C-Test: Beiträge aus der aktuellen Forschung / The C-test: Contributions from current research*, S. 125–192. Frankfurt: Lang.

Eckes, T. (2010b): Rasch models for C-tests: Closing the gap on modern psychometric theory. In: A. Berndt & K. Kleppin (Hrsg.): *Sprachlehrforschung: Theorie und Empirie – Festschrift für R. Grotjahn*, S. 39–49. Frankfurt: Lang.

Eckes, T. (2010c): Standard-Setting bei C-Tests: Bestimmung von Kompetenzniveaus mit der Prototypgruppenmethode. *Diagnostica*, 56, S. 19–32.

Eckes, T. (2011a): Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments. Frankfurt: Lang.

Eckes, T. (2011b): Item banking for C-tests: A polytomous Rasch modeling approach. *Psychological Test and Assessment Modeling*, 53, S. 414–439.

Edelhoff, C. (2010): Kommunikative Kompetenz revisited – Anmerkungen in einer überflüssigen Debatte, *ForumSprache*, 4, S. 149–160.

Ehlich, K. (Hrsg.) (1996): *Kindliche Sprachentwicklung: Konzepte und Empirie*. Opladen: Westdeutscher Verlag.

Ehlich, K. (2007): Sprachaneignung und deren Festlegung bei Kindern mit und ohne Migrationshintergrund – Was man weiß, was man braucht, was man erwarten kann. In: K. Ehlich et al. (2007): Anforderungen an Verfahren der regelmäßigen Sprachstandsfeststellung als Grundlage für die frühe und individuelle Förderung von Kindern mit und ohne Migrationshintergrund, Bd. 11, S. 11-75, Bildungsforschung. Berlin/Bonn: Bundesministerium für Bildung und Forschung.

Ehlich, K. (2012): Sprach(en)aneignung – mehr als Vokabeln und Sätze. Stiftung Mercator, proDaZ Deutsch, Universität Duisburg-Essen, Oktober 2012, http://www.uni-due.de/imperia/md/content/prodaz/sprach_en_aneignung_mehr_als_vokabeln_und_s_ätze.pdf.

Eichler, W. (2007): Sprachbewusstheit Deutsch. In: B. Beck. & E. Klieme (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie (Deutsch Englisch Schülerleistungen International), S. 112-120, Weinheim: Beltz.

Erpenbeck, J. & Heyse, V. (1999): Die Kompetenzbiografie. Strategien der Kompetenzentwicklung durch selbstorganisiertes Lernen und multimediale Kommunikation. Münster [u. a.]: Waxmann.

Erpenbeck, J. & Rosenstiehl, L. v. (Hrsg.) (2003): Handbuch Kompetenzmessung. Stuttgart Schäffer-Poeschel.

Feilke, H. & Augst, G. (1993): Schreiben, Schreibschwächen und Grammatik in der Schule. Oder: Der gewendete „rote Pollover“. In: Der Deutschunterricht 2, S. 90-96 (Wiederabdruck in DU 2/1994).

Feldmann, U., Grotjahn, R. & Stemmer, B. (1986): Was messen Sprachtests eigentlich? In: Seminar für Sprachlehrforschung der Ruhr-Universität Bochum (Hrsg.): Probleme und Perspektiven der Sprachlehrforschung. Bochumer Beiträge zum Fremdsprachenunterricht in Forschung und Lehre, S. 325-338. Frankfurt a. M.: Scriptor.

Freese, H.-L. (1994): Was mißt und was leistet 'Leistungsmessung mittels C-Tests'? In: R. Grotjahn (Hrsg.) (1994): S. 305-311. [Slightly revised version of an unpublished paper (1985) with the same title].

Glück, H. (Hrsg.) (2000): Metzler Lexikon Sprache, 2. Aufl., Stuttgart/Weimar: Metzler.

Gogolin, I. & Neumann, U. (Hrsg.) (2009): Streitfall Zweisprachigkeit - The Bilingualism Controversy. Wiesbaden: VS-Verlag.

Goodman, K. S. (1967): Reading: a Psycholinguistic Guessing Game. In: Journal of the Reading Specialist, 6, S. 126-135.

Goodman, K. S. (Hrsg.) (1973): *The Psycholinguistic Nature of the Reading Process*. Wayne: State University Press.

Grimm, H. (1999): *Störungen der Sprachentwicklung. Grundlagen – Ursachen – Diagnose – Intervention – Prävention*. Göttingen: Hogrefe.

Groeben, N. (1978): *Die Verständlichkeit von Unterrichtstexten*. Münster: Aschendorff.

Groeben, N. & Hurrelmann, B. (2002): *Lesekompetenz: Bedingungen, Dimensionen, Funktionen*. Weinheim: Juventa.

Groß Ophoff; J., Isaac; K., Hosenfeld; I. & Eichler, W. (2008): Erfassung von Leseverständnis im Projekt VERA. In: B. Hofmann & R. Valtin (Hrsg.): *Checkpoint Literacy - Tagungsband 2 zum 15. Europäischen Lesekongress 2007 in Berlin*, S. 36-51. Berlin: Deutsche Gesellschaft für Lesen und Schreiben.

Grotjahn, R. (1984): Ist der C-Test ein Lesetest? In: A. Addison & K. Vogel (Hrsg.): *Lehren und Lernen von Fremdsprachen im Studium*, S. 230-248. Bochum: AKS.

Grotjahn, R. (1987a): How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In: R. Grotjahn; C. Klein-Braley & D. K. Stevenson (Eds.): *Taking their measure: The validity and validation of language tests*, p. 219-253. Bochum: Brockmeyer.

Grotjahn, R. (1987b): Ist der C-Test ein Lesetest? In: A. Addison & K. Vogel (Hrsg.): *Lehren und Lernen von Fremdsprachen im Studium*, S. 230-248. Bochum: AKS.

Grotjahn, R. & Tönshoff, W. (1992): Textverständnis bei der C-Test-Bearbeitung. Pilotstudien mit Französisch- und Italienischlernern. In: R. Grotjahn (Ed.) (1992): *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*, Bd. 1, S. 19-95. Bochum: Brockmeyer.

Grotjahn, R.; Klein-Braley, C. & Raatz, U. (1992): C-Tests in der praktischen Anwendung. Erfahrungen beim Bundeswettbewerb Fremdsprachen. In: R. Grotjahn (Hrsg.): *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*, Bd. 1, S. 263-296. Bochum: Brockmeyer.

Grotjahn, R. & Stemmer, B. (1996): C-Tests and language processing. In: J. A. Coleman; R. Grotjahn & U. Raatz (Eds.): *The European Language Proficiency Survey: an overview of findings*. University language learning and the C-Test, p. 115-130. Bochum: AKS-Verlag.

Grotjahn, R. (1997): Der C-Test: Neuere Entwicklungen. In: M. Gardenghi & M. O'Connell (Hrsg.): *Prüfen, Testen, Bewerten im modernen Fremdsprachenunterricht*, S. 117-128. Frankfurt am Main: Lang.

Grotjahn, R. (2000): Sprachtests: Formen und Funktionen. In: Hallet, W. & Königs, F. G. (Hrsg.): Handbuch Fremdsprachendidaktik, S. 211-215. Seelze-Velbert: Kallmeyer.

Grotjahn, R.; Klein-Braley, C. & Raatz, U. (2002): C-Tests: an Overview. In: J. A. Coleman, R. Grotjahn & U. Raatz (Eds.): University Language Testing and the C-Test, S. 93-114. Bochum: AKS-Verlag.

Grotjahn, R. (2003): Leistungsmessung und Leistungsbewertung. Fernstudienbrief für den Weiterbildungs-Masterstudiengang „Deutschlandstudien. Schwerpunkt: Deutsche Sprache und ihre Vermittlung“. FernUniversität – Gesamthochschule in Hagen.

Grotjahn, R. (2009): Testen im Fremdsprachenunterricht: Aspekte der Qualitätsentwicklung. Praxis Fremdsprachenunterricht 1, S. 4-8.

Grotjahn, R.; Schlak, T. & Aguado, K. (2010): S-C-Tests: Messung automatisierter sprachlicher Kompetenzen anhand von C-Tests mit massiver textspezifischer Zeitlimitierung. In: R. Grotjahn (Hrsg.): Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from current research, S. 297-319. Frankfurt am Main: Lang.

Grotjahn, R. (2010a): Gesamtdarbietung, Einzeltextdarbietung, Zeitbegrenzung und Zeitdruck: Auswirkungen auf Item- und Testkennwerte und C-Test-Konstrukt. In: R. Grotjahn (Hrsg.): Der C-Test: Beiträge aus der aktuellen Forschung/The C-Test: Contributions from current research, S. 265-296. Frankfurt/Main: Lang.

Grotjahn, R. (2010b): Sprachtests: Formen und Funktionen. In: W. Hallet & F. G. Königs (Hrsg.): Handbuch Fremdsprachendidaktik, S. 211-215. Seelze-Velber: Kallmeyer.

Grotjahn, R. (2011): C-Tests – Aspekte der Validität. Deutsch als Fremdsprache, 48 (3), S. 131-137.

Gummich, V. (1997): C-Test-Leistung, Schultyp, Schulstufe. Gerhard-Mercator-Universität-GH-Duisburg, Fachbereich 2: Psychologie. Unveröffentlichte Diplomarbeit.

Guthrie, J. T.; Wigfield, A.; Metsala, J. L. & Cox, K. E. (1999): Motivational and cognitive predictors of text comprehension and reading amount. Scientific Studies of Reading, 3, p. 231-256.

Harris, D. P. (1969): Testing English as A Second Language. New York: McGraw-Hill.

Harsch, C. & Schröder, K. (2007): Textrekonstruktion Englisch. In: B. Beck. & E. Klieme (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie (Deutsch Englisch Schülerleistungen International), S. 149-157, Weinheim: Beltz.

Harsch, C.; Neumann, A.; Lehmann, R. & Schröder, K. (2007): „Schreibfähigkeit“. In: B. Beck & E. Klieme (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie, S. 42-62. Weinheim: Beltz.

Harsch, C. & Schröder, K. (2008): „Textrekonstruktion Englisch“. In: DESI Konsortium (Hrsg.): Unterricht und Kompetenzerwerb in Deutsch und Englisch, S. 139-148. Weinheim: Beltz.

Harsch, C. & Hartig, J. (2010): Empirische und inhaltliche Analyse lokaler Abhängigkeiten im C-Test. In: R. Grotjahn (Hrsg.): Der C-Test: Beiträge aus der aktuellen Forschung, S. 193- 204. Frankfurt a. M.: Peter Lang.

Hartig, M. (1999): Psycholinguistik des Deutschen, Germanistische Lehrbuchsammlung, Bd. 17. Berlin: Weidler.

Hartig, J. (2007): Skalierung und Definition von Kompetenzniveaus. In: B. Beck. & E. Klieme (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie (Deutsch Englisch Schülerleistungen International), S. 83-99, Weinheim: Beltz.

Hartig, J. & Jude, N. (2008): Sprachkompetenzen von Mädchen und Jungen. In: DESI-Konsortium [Hrsg.]: Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie, S. 202-207. Weinheim: Beltz.

Hastings, A. J. (2002): Error analysis of an English C-test: Evidence for integrated processing. In: Grotjahn, R. (Hrsg.): Der C-Test. Theoretische Grundlagen und praktische Anwendungen, Bd. 4, S. 53-66. Bochum: AKS-Verlag.

Heinemann, M. & Höpfner, C. (1993): Screeningverfahren zur Erfassung von Sprachentwicklungsverzögerungen SEV. Weinheim: Beltz.

Helmke, A. & Hosenfeld, I. (2003a): Vergleichsarbeiten (VERA): Eine Standortbestimmung zur Sicherung schulischer Kompetenzen - Teil 1: Grundlagen, Ziele, Realisierung. SchulVerwaltung, Ausgabe Hessen/Rheinland-Pfalz/Saarland, 1, S. 10-13.

Helmke, A. & Hosenfeld, I. (2003b): Vergleichsarbeiten (VERA): Eine Standortbestimmung zur Sicherung schulischer Kompetenzen - Teil 2: Nutzung für Qualitätssicherung und Verbesserung der Unterrichtsqualität. SchulVerwaltung, Ausgabe Hessen/Rheinland-Pfalz/Saarland, 2, S. 41-43. [ebenfalls in: Ausgabe Nordrhein-Westfalen, 5, S. 143-145.]

Hufeisen, B. (1997): Rezension. Stemmer Brigitte. What's on a C-test taker's mind? Mental processes in C-test taking. Bochum: Universitätsverlag

Brockmeyer, 1991 (Manuskripte zur Sprachlehrforschung 36). Zeitschrift für interkulturellen Fremdsprachenunterricht, <http://zif.spz.tu-darmstadt.de/jg-02-2/beitrag/stemmer.htm>.

Hurrelmann, B. (2002): Prototypische Merkmale der Lesekompetenz. In: N. Groeben & B. Hurrelmann (Hrsg.): Lesekompetenz. Bedingungen, Dimensionen, Funktionen, S. 275-286. Weinheim: Juventa.

Hymes, D. (1971): Competence and performance in linguistic theory. In: Huxley, R. & Ingram, E. (Hrsg.): Language Acquisition and Methods. New York: Academic Press.

Ingenkamp, K. (1999): Die Fragwürdigkeit der Zensurenggebung. Texte und Untersuchungsberichte. Weinheim: Beltz.

Isaac, K. (2010a): Diagnostisches Potenzial von Vergleichsarbeiten. In: S. Kliemann (Hrsg.): Diagnostizieren und Fördern - Kompetenzen erkennen, unterstützen und erweitern. Beispiele und Anregungen für die Jahrgänge 1 bis 4, S. 59-65. Berlin: Cornelsen Verlag Scriptor.

Isaac, K. (2010b): Chancen und Grenzen von Vergleichsarbeiten - Zur schulinternen Nutzung von Daten. Schulverwaltung NRW, 21, 7/8, S. 197-199.

Isaac, K.; Hochweber, J. & Eichler, W. (2010): Differenzierende Modellierung der Aufgabenschwierigkeit im Bereich „Sprache und Sprachgebrauch untersuchen“. In: B. Schwarz, P. Nenniger & R. S. Jäger (Hrsg.): Erziehungswissenschaftliche Forschung - nachhaltige Bildung. Beiträge zur 5. DGfE-Sektionstagung "Empirische Bildungsforschung"/AEPFKBBB im Frühjahr 2009. Landau: Verlag Empirische Pädagogik.

Jafarpur, A. (2002): A comparative study of the C-test and a cloze test. In: R. Grotjahn (Hrsg.): Der C-Tests. Theoretische Grundlagen und praktische Anwendungen, Bd. 4, p. 31-51. Bochum: AKS-Verlag.

Jakschik, G. (1992): Zum Einsatz des C-Tests in den Psychologischen Diensten der Arbeitsämter. Ein C-Test für Deutsch als Zweitsprache. In: R. Grotjahn (Hrsg.): Grotjahn, Rüdiger (Hrsg.): Der C-Test. Theoretische Grundlagen und praktische Anwendungen, Bd. 1, S. 297-311. Bochum: Brockmeyer.

Jakschik, G. (1996): Validierung des C-Tests für erwachsene Zweitsprachler. Eine Längsschnittuntersuchung bei Trägern von beruflichen Bildungsmaßnahmen. In: R. Grotjahn (Ed.) (1996): Der C-Test. Theoretische Grundlagen und praktische Anwendungen (Vol. 3), S. 235-277. Bochum: Brockmeyer.

Johansson, S. (1973): Partial dictation as a test of foreign language proficiency, Swedish-English Contrastive Studies Report Number 3, Department of English: Lund University

Jonz, J. (1976): Improving on the basic egg: the multiple-choice cloze. *Language Learning*, 26, 2, p. 255-265.

Jude, N. & Klieme, E. (2007): Sprachliche Kompetenz aus Sicht der pädagogisch-psychologischen Diagnostik. In: B. Beck. & E. Klieme (Hrsg.): *Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie (Deutsch Englisch Schülerleistungen International)*, S. 9-22. Weinheim: Beltz.

Jude, N. (2008): Zur Struktur von Sprachkompetenz, http://www.fachportal-paedagogik.de/fis_bildung/suche/fis_set.html?FId=880590.

Kanning, U. P. (2003): *Diagnostik sozialer Kompetenzen*. Göttingen [u.a.]: Hogrefe.

Kaufhold, M. (2006): *Kompetenz- und Kompetenzerfassung. Analyse und Beurteilung von Verfahren der Kompetenzerfassung*. Wiesbaden: GWV Fachverlage GmbH.

Kecker, G. (2011): *Validierung von Sprachprüfungen. Die Zuordnung des TestDaF zum Gemeinsamen europäischen Referenzrahmen für Sprachen*. Frankfurt am Main [u.a.]: Lang.

Khalifa, H. & Weir, C. (2009): *Examining Reading: Research and Practice in Assessing Second Language Reading*, *Studies in Language Testing*, 29. Cambridge: UCLES/Cambridge University Press.

Kintsch, W. (1992): A cognitive architecture for comprehension. In: H. L. Pick; P. van Broek & D. C. Knill (Eds.): *The study of cognition: Conceptual and methodical issues*, p. 143-164. Washington, DC: American Psychological Association.

Klein, W. (2000): Prozess des Zweitspracherwerbs. In: H. Grimm (Hrsg.): *Sprachentwicklung, Enzyklopädie der Psychologie*, Bd. C/III/3, S. 537-570, Göttingen: Hogrefe.

Klein-Braley, C. & Raatz, U. (1984): A survey of research on the C-Test. *Language Testing*, 1, p. 134-146.

Klein-Braley, C. & Raatz, U. (1985): C-Test as placement tests for German university students of English. In: C. Klein-Braley & U. Raatz (Hrsg.): *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis*, S. 96-100. Bochum: AKS.

Klein-Braley, C. & Raatz, U. (1985a): C-Tests and construct validity. In: C. Klein-Braley & U. Raatz (Hrsg.): *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis*, S. 55-65. Bochum: AKS.

Klein-Braley, C. & Raatz, U. (1985b): Reduced redundancy as an approach to language testing. In: C. Klein-Braley & U. Raatz (Hrsg.): Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis, S. 1-13, Bochum: AKS.

Klein-Braley, C. & Raatz, U. (1985c): Tests of reduced redundancy – theory. In: V. Kohonen & A. J. Pitkänen (Eds.): Language Testing in School. AfinLA Yearbook 1985, S. 33-48. Tampere: AfinLA.

Klein-Braley, C. (1996): Towards a theory of C-Test processing. In: R. Grotjahn (Hrsg.): Der C-Test. Theoretische Grundlagen und praktische Anwendungen, Bd. 3. Bochum: Brockmeyer.

Klein-Braley, C. (1997): C-Tests in the context of reduced redundancy testing: an appraisal. Language Testing, 14, 1, S. 47-84.

Klein-Braley, C. (1981): Empirical investigations of cloze tests. An examination of the validity of cloze tests as tests of general language proficiency in English for German university students. Universität-Gesamthochschule-Duisburg, Fachbereich 3: Sprach- und Literaturwissenschaften. Unveröffentlichte Dissertation.

Klein-Braley, C. & Raatz, U. (1982): Der C-Test: ein neuer Ansatz zur Messung von allgemeiner Sprachbeherrschung, S. 23-37. Bochum: AKS

Klein-Braley, C. (1984): Advanced prediction of difficulty with C-Tests. In: T. Culhane; C. Klein-Braley & D. K. Stevenson (Eds): Practice and problems in language testing 7. Proceedings of the Sevenths International Language Testing Symposium of the Interuniversitärer Sprachtestgruppe (IUS). Colchester: University of Essen.

Klein-Braley, C. (1985a): Reduced redundancy as an approach to language testing. In: Fremdsprachen und Hochschule, S. 1-13, Bochum: AKS.

Klein-Braley, C. (1985b): Advance prediction of test difficulty. In: C. Klein-Braley & U. Raatz (Eds.): Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis, S. 1-13. Bochum: AKS.

Klein-Braley, C. (1994): Language testing with the C-test. A linguistic and statistical investigation in to the strategies used by C-test takers, and the prediction of C-test difficulty. Universität-Gesamthochschule-Duisburg, Fachbereich 3: Sprach- und Literaturwissenschaften: Unveröffentlichte Habilitationsschrift.

Klein-Braley, C. (1987): "Fossil it large: translation as a language testing procedure." In: R. Grotjahn; C. Klein-Braley & D. K. Stevenson (Eds.): Their Measure: The Validity and Validation of Language Tests, p. 111-132. Bochum: Brockmeyer.

Klieme et al. (2003): Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Koordination: Deutsches Institut für Internationale Pädagogische Forschung (DIPF), verfügbar unter:

http://www.dipf.de/publikationen/volltexte/zur_entwicklung_nationaler_bildungsstandards.pdf.

Klieme, E.; Eichler, W.; Helmke, A.; Lehmann, R.; Nold, G.; Rolff, H.-G.; Schröder, K.; Thomé, G. & Willenberg, H. (2006): Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch-Englisch-Schülerleistungen-International (DESI). Deutsches Institut für Internationale Pädagogische Forschung. Frankfurt am Main, <http://www.dipf.de/de/projekte/pdf/biqua/desi-zentrale-befunde>.

Kintsch, W. (1994): Text comprehension, memory, and learning. *American Psychologist*, 49, p. 294-303.

Kintsch, W. (1998): *Comprehension: A Paradigm for Cognition*. Cambridge: Cambridge University Press.

Krekeler, C. (2002): DSH und TestDaF - zwei ungleiche Sprachtests im Vergleich. *ELiSe: Essener Linguistische Skripte - elektronisch* 2, 2, S. 19-50. Neuabdruck als: Krekeler, C. (2004): TestDaF und DSH - ungleiche Sprachtests im Vergleich. In: H. Casper-Hehne & U. Koreik: *DSH und TestDaF als hochschulbezogene Prüfungssysteme für Deutsch als Fremdsprache. Standortbestimmungen und Entwicklungslinien (= Perspektiven Deutsch als Fremdsprache)*, Bd. 17, S. 134-165. Baltmannsweiler: Schneider Verlag Hohengehren.

Krekeler, C. (2005): *Grammatik und Fachbezug in Sprachtests für den Hochschulzugang*. Dissertation. Universität Essen.

Kultusministerkonferenz (KMK) (2003): Beschlüsse der Kultusministerkonferenz. Bildungsstandards im Fach Deutsch für den mittleren Schulabschluss, Beschluss vom 4.12.2003. Luchterhand,

http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-BS-Deutsch-MS.pdf.

Lado, R. (1961): *Language Testing: The Construction and Use of Foreign Language Tests*. New York: McGraw-Hill Book Company.

Lehmann, R. H.; Peek, R. & Poerschke, J. (1997): *HAMLET 3-4. Hamburger Lesetest für 3. und 4. Klassen*. Beltz: Weinheim.

Lehmann, R. H; Gänsfuß, R. & Peek, R. (1999): *Aspekte der Lernausgangslage der Lernentwicklung – Klassenstufe 7*. Hamburg: Behörde für Schule, Jugend und Berufsbildung.

Lempert, W. (1998): Berufliche Sozialisation oder was Berufe aus Menschen machen. Grundlagen der Berufs- und Erwachsenenbildung, Bd. 16. Baltmannsweiler. Schneider Verlag Hohendehren.

Lenhard, W. & Schneider, W. (Hrsg.) (2009): Diagnostik und Förderung des Leseverständnisses. Göttingen: Hogrefe.

Lienert, G. A. & Raatz, U. (1998): Testbau und Testanalyse. Weinheim: Beltz.

Linke, A.; Nussbaumer, M. & Portmann P. R. (2001): Studienbuch Linguistik, Tübingen: Niemeyer.

Lisker, A. (2010): Sprachstandsfeststellung und Sprachförderung im Kindergarten sowie beim Übergang in die Schule. München: DJI.

Lutjeharms, M. (2010): Der Leseprozess in Mutter- und Fremdsprache. In: M. Lutjeharms & C. Schmidt: Lesekompetenz in Erst-, Zweit- und Fremdsprache. Tübingen: Narr.

Marx, H. (1998): Knuspels-Leseaufgaben (KNUSPEL-L): Handanweisung, Göttingen [u.a.]: Hogrefe.

May, P. (1996): Die Hamburger Schreibprobe. Grundlegende Rechtschreibstrategien erfassen. Grundschule, 4, S. 17 - 20.

May, P. (1999): Strategiebezogene Rechtschreibdiagnose - mit und ohne Test: Analyse von freien Schreibungen mit Hilfe der HSP-Kategorien. In: H. Balhorn; H. Bartnitzky; I. Büchner & A. Speck-Hamdan (Hrsg.): Schatzkiste Sprache I: Lesen und Schreiben von Anfang an. AKG-Band 103. Frankfurt a. M.: Arbeitskreis Grundschule.

May, P. & Malitzky, V. (1999): Erfassung der Rechtschreibkompetenz in der Sekundarstufe mit der Hamburger Schreibprobe (HSP 4/5 und HSP 5-9). In: Lade, E. & Kowalczyk, W. (Hrsg.): Konkrete Handlungsanleitungen für erfolgreiche Beratungsarbeit mit Schülern, Eltern und Lehrern. Kissing: WEKA Fachverlag.

May, P.; Malitzky, V. & Vieluf, U. (2001): Rechtschreibtests im Vergleich: Wie stellt man deren Güte fest und wie besser nicht? Anmerkungen zur Kritik von Tacke, Völker und Lohmüller an der HSP. Psychologie in Erziehung und Unterricht, 2, S. 146-152.

McKeown, M. G.; Beck, I. L.; Omanson, R. C. & Perfetti, C. A. (1987): The effects of long-term vocabulary instruction on reading comprehension: A replication. Journal of Reading Behavior, 15, p. 3-18.

Metze, W. (2009): Stolperwörter-Lesetest. Handanweisungen, Version 2009, http://wilfriedmetze.de/Handanweisung_2009.pdf.

Moore, A. & Theunissen, A. F. (1994): Qualifikation versus Kompetenz. Eine Diskussion um Begriffe, die Entwicklung neuer Modelle oder eine Frage von politischem Rang? Europäische Zeitschrift für Berufsbildung, 1, S. 74-80.

Moravcsik, J. E. & Kintsch, W. (1993): Writing quality, reading skill, and knowledge as factors in text comprehension. Canadian Journal of Experimental Psychology, 47, p. 360-374.

Morrow, K. (1979): Communicative language testing: revolution or evolution? In: Brumfit, C. J. & Johnson, K. (Eds.). The Communicative Approach to Language Teaching. Oxford: Oxford University Press.

Nodari, C. (2002): Was heisst eigentlich Sprachkompetenz? In: Barriere Sprachkompetenz. Dokumentation zur Impulstagung vom 2.11.2001 im Volkshaus Zürich, SIBP Schriftenreihe, 18, S. 9-14.

Nold, G. & Willenberg, H. (2007): Lesefähigkeit [Englisch und Deutsch]. In: B. Beck. & E. Klieme (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messungen. DESI-Studie (Deutsch Englisch Schülerleistungen International), S. 23-41, Weinheim: Beltz.

North, B. (2000): The development of a common framework scale of language proficiency (Theoretical studies in second language acquisition), Bd. 8. New York: Lang.

OECD (2003): OECD-Veröffentlichung „Bildung auf einen Blick 2003“. Wesentliche Aussagen der OECD zur Ausgabe 2003, 16.09.2003, http://www.bmbf.de/pub/20030916_eag_langfassung.pdf.

OECD (2004): Education at Glance 2004, <http://www.oecd.org/edu/skills-beyond-school/educationataglance2004-home.htm>.

OECD (2011): Bildung auf einen Blick 2011: OECD-Indikatoren, <http://www.oecd-ilibrary.org/docserver/download/9611045e.pdf?expires=1373969819&id=id&accname=ocid57016100&checksum=19FD8CE366B015202AA381BDECE85614>.

Oller, J.W. Jr. & Nevin, N. (1971a): A cloze test of English prepositions. TESOL Quarterly, 5, p. 315-326.

Oller, J. W. Jr. (1971b): Dictation as a device for testing foreign language proficiency. ELT 25, p. 254-259.

Oller, J. W. Jr. (1972): Scoring methods and difficulty levels for cloze tests of proficiency in ESL. Modern Learning Journal, 56, p. 151-158.

Oller, J. W. Jr. (1973): Cloze tests of second language proficiency and what they measure. Language learning, 23, p. 105-118.

Oller, J. W. Jr. (1979): Language tests at school. London: Longman.

Oller, J. W. Jr. (1983): Evidence for a general language proficiency factor: an expectancy grammar. In: J.W. Jr.: Issues in language testing research, p. 3-10. Rowley, Mass.: Newbury House.

Oller, J. W. Jr. (1986): Language tests at School. London: Longman.

Oller, J. W. Jr. & Hinofotis, F. (1978): Two mutually exclusive hypotheses about second language ability: factor analytic studies of a variety of language tests. In: J.W. Jr. Oller & K. Perkins (Eds.): Language in education: testing the tests. Rowly, Mass.: Newbury House.

Oller, J. W. (1991): Foreign Language Testing, 1: Its Breadth, ADFL Bulletin, 22, p. 33-38.

Pietrzyk, U. (2002): Brüche in der Berufsbiografie – Chancen und Risiken für die Entwicklung beruflicher Kompetenz. Hamburg: Kovac.

Pohl, T. (2007): Studien zur Ontogenese wissenschaftlichen Schreibens, Reihe Germanistische Linguistik, 271. Tübingen: Niemeyer.

Pospiech, U. (2005): Schreibend schreiben lernen. Über die Schreibhandlung zum Text als Sprachwerk. Frankfurt/Main: Peter Lang. [auch unter www.elise.uni-essen.de, Beiheft 1/2004. Download: http://www.uni-due.de/germanistik/elise/beiheft_12004.

Raatz, U. & Klein-Braley, C. (1983): Ein neuer Ansatz zur Messung der Sprachleistung. Der C-Test: Theorie und Praxis. In: R. Horn; K. Ingenkamp & R. S. Jäger (Hrsg.): Tests und Trends 3. Jahrbuch der Pädagogischen Diagnostik, S. 107-138. Weinheim & Basel: Beltz.

Raatz, U. (1985): Tests of reduced redundancy - the C-test, a practical example. Fremdsprachen und Hochschulen, S. 14-19. Bochum: AKS.

Raatz, U. & Klein-Braley, C. (1985a): Ein neuer Ansatz zur Messung der Sprachleistung. Der C-Test: Theorie und Praxis. In: R. Horn; K. Ingenkamp; R. Jäger (Hrsg.): Tests und Trends: Jahrbuch der Pädagogischen Diagnostik, S. 107-138. Weinheim: Beltz.

Raatz, U. & Klein-Braley, C. (1985b): How to develop a c-test. In: Fremdsprachen und Hochschule, 13/14, S. 20-22. Bochum: AKS.

Raatz, U.; Voss, B. & Klein-Braley, C. (1991): Diagnose der Fremdsprachenleistung in der Schule. In: K. Ingenkamp & R. Jäger (Hrsg.): Tests und Trends: Jahrbuch der Pädagogischen Diagnostik.

Raatz, U. & Klein-Braley, C. (1994): Analyse der Ergebnisse im Einzelwettbewerb des Bundeswettbewerbs Fremdsprachen: Wettbewerb Sekundarstufe I Frühjahr 1991. In: R. Grotjahn (Hrsg.): Der C-Test. Theoretische Grundlagen und praktische Anwendungen, Bd. 2, S. 239-258. Bochum: Brockmeyer.

Reitbauer, M. (2000): Effektiver lesen mit Superstrukturen: eine empirische Untersuchung anhand der Fachtextsorte Abstract. Tübingen: Narr.

Rey, G. D. (2012): Methoden der Entwicklungspsychologie. Datenerhebung und Datenauswertung. Norderstedt bei Hamburg: BoD, <http://www.methoden-psychologie.de/index.html>.

Richter, T. & Christmann, U. (2002): Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In: N. Groeben & B. Hurrelmann (Hrsg.): Lesekompetenz: Bedingungen, Dimensionen, Funktionen, S. 25-58. Weinheim: Juventa.

Rosebrock, C. (2007): Reading Literacy und Lesekompetenz. Die kognitive Dimension des Lesens und die innere Beteiligung des Lesers, Artikel vom 23.11.2007, verfügbar unter: <http://www.lesen-in-deutschland.de/html/content.php?object=journal&lid=778>.

Rosebrock, C. & Nix, D. (2012): Grundlagen der Lesedidaktik und der systematischen schulischen Leseförderung. Baltmannsweiler: Schneider Verlag Hohengehren.

Roth, H.-J. (2008): Verfahren zur Sprachstandsfeststellung – ein kritischer Überblick. In: C. Bainski & M. Krüger-Potratz (Hrsg.): Handbuch Sprachförderung, S. 22-41 Essen: Neue Deutsche Schule.

Sang, F. & Vollmer, H. J. (1980): Modelle linguistischer Kompetenz und ihre empirische Fundierung. In: R. Grotjahn. & E. Hopkins (Eds.): Empirical Research on Language Teaching and Language Acquisition (Quantitative linguistics), Bd. 6, S. 3-80. Bochum: Brockmeyer.

Schiefele, U. (1996): Motivation und Lernen mit Texten. Göttingen: Hogrefe.

Schindler, K. & Siebert-Ott, G. (2010): Schreiben in der Zweitsprache. In: H. Feilke & T. Pohl (Hrsg.): Schriftlicher Sprachgebrauch/Texte verfassen. (Reihe Deutschunterricht in Theorie und Praxis, herausgegeben von Winfried Ulrich). Baltmannsweiler: Schneider Verlag Hohengehren.

Schmidt-Atzert, L. & Amelang, M. (2006): Psychologische Diagnostik: mit 82 Tabellen. Berlin: Springer.

Scholten-Akoun, D. & Baur, R. S. (unter Mitarbeit von A. Mashkovskaya) (2012): Der C-Test als ein Instrument zur Messung der Schriftsprachkompetenzen von Lehramtsstudierenden (auch) mit Migrationshintergrund – eine Studie. In: B. Ahrenholz & W. Knapp (Hrsg.) (2012): Sprachstand erheben – Spracherwerb

erforschen. Beiträge aus dem 6. Workshop „Kinder mit Migrationshintergrund“, 2010, S. 307-330. Freiburg i. Br.: Fillibach.

Scholten-Akoun, D.; Tischmeyer, D. & Mashkovskaya, A. (2013, in Vorbereitung): Students' language competencies. Design and results of an empirical study. Results from the International workshop from 04.10.2012 till 06.20.2012 in Wuppertal, "heritage languages: language contact-change-maintenance and loss in the wave of new migration landscapes".

Scholten-Akoun, D.; Kuhnen, A. & Mashkovskaya, A. (2012): Sprachkompetenzen Studierender. Design und erste Ergebnisse einer empirischen Studie. In: H. Feilke; J. Köster & M. Steinmetz (Hrsg.) (2012): Textkompetenzen in der Sekundarstufe II, S. 207-228. Freiburg i. Br.: Fillibach.

Schröder, K. (2007): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International), S. 290-298. Weinheim u.a. : Beltz.

Shannon, C. E. (1948): A Mathematical Theory of Communication. Reprinted with corrections from The Bell System Technical Journal, 27, p. 379-423, 623-656, July, October, 1948,
<http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>.

Shannon, C. E. & Weaver, W. (1949): The mathematical theory of communication. Urbana: University of Illinois Press.

Sigott, G. & Köberl, J. (1993): Validating the X-test. Language Testing Update, 14, p. 53-58.

Sigott, G. (1998): C-Testing in 1998: A Résumé. Referat anlässlich der internationalen Konferenz: Testen/Prüfen, Leistungsbeurteilung und Evaluation. Wien (18.-25. März).

Sigott, G. (2002): "High-level process in C-Test taking?" In: R. Grotjahn (Hrsg.): Der C-Test. Theoretische Grundlagen und praktische Anwendungen, Bd. 4, S. 67-82. Bochum: AKS-Verlag.

Sigott, G. (2004): Towards identifying the C-Test construct. Frankfurt am Main, New York: Peter Lang.

Souvignier, E. (2009): Leseförderung: Diagnose der Lesekompetenz. Schule NRW, 11, S. 545-547.

Spolsky, B. (1973): What does it mean to know a language; or how do you get somebody to perform his competence? In: J. W. Jr. Oller & J. Richards (Eds.): Focus on the learner, p. 164-176. Rowley, Mass.: Newbury House.

Spolsky, B. et al. (1968): Preliminary studies in the development of techniques for testing overall second language proficiency. In: J. A. Upshur & J. Fata (Eds.): Problems in foreign language testing. Language Learning Special Issue No. 3: p. 79-103.

Spolsky, B. (1995): Measured words: the development of objective language testing. Oxford: Oxford University Press.

Steinhoff, T. (2007): Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten. Tübingen: Niemeyer.

Stemmer, B. (1991): What's on a C-test taker's mind? Mental processes in C-test taking. Manuskripte zur Sprachlehrforschung, Bd. 36. Bochum: Brockmeyer.

Süßmilch, E. (1984a): Sprachleistungsmessung mittels C-Tests. Arbeitsberichte aus der Abteilung für Bildungsforschung und Pädagogische Beratung des Erziehungswissenschaftlichen Instituts der Universität Düsseldorf, 1. [also published as Süßmilch, 1984b].

Süßmilch, E. (1984b): Sprachleistungsmessung mittels C-Tests. Finlance, 3, S. 55-93.

Szagun, G. (2006): Sprachentwicklung beim Kind: Ein Lehrbuch. Weinheim: Beltz.

Taylor, W. L. (1953): Cloze procedure: a new tool for measuring readability. Journalism Quarterly, p. 415-433.

TestDaF-Institut (2009) (Hrsg.): Muster-Prüfungen. Ismaning: Max Hueber Verlag.

Thonke, F.; Groß Ophoff, J., Hosenfeld, I. & Isaac, K. (2008): Kriteriengestützte Erfassung von Schreibleistungen im Projekt VERA. In: B. Hofmann & R. Valtin (Hrsg.): Checkpoint Literacy - Tagungsband 2 zum 15. Europäischen Lesekongress 2007 in Berlin, S. 28-35. Berlin: Deutsche Gesellschaft für Lesen und Schreiben.

Trautner, H. M. (1991): Lehrbuch der Entwicklungspsychologie. Göttingen: Hogrefe.

Trim, J.; North, B., Coste, D. & Sheils, J. (2001): Gemeinsamer Europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin, München: Langenscheidt KG,
<http://student.unifr.ch/pluriling/assets/files/Referenzrahmen2001.pdf>.

van Dijk, T. A. & Kintsch, W. (1983): Strategies of discourse comprehension. New York: Academic Press.

Vollmer, H. J. (1982): Spracherwerb und Sprachbeherrschung: Untersuchungen zur Struktur von Fremdsprachenfähigkeit. Tübingen: Narr.

Weinert, F. E. (1996): Lerntheorien und Instruktionsmodelle. In: F. E. Weinert (Hrsg.): Enzyklopädie der Psychologie. Pädagogische Psychologie, Bd. 2, Psychologie des Lernens und der Instruktion, S. 1-48. Göttingen: Hogrefe.

Weinert, F. E. (2001a): Concept of competence. A conceptual classification. In: D. S. Rychen & L. H. Salganik (Eds.): Defining and selecting key competencies, p. 45-66. Seattle: Hogrefe & Huber.

Weinert, F. E. (2001b): Vergleichende Leistungsmessung in Schulen – ein umstrittenes Selbstverständnis. In: F. E. Weinert (Hrsg.): Leistungsmessung in Schulen, S. 17-32. Weinheim: Beltz.

Weiß, R. (1999): Erfassung und Bewertung von Kompetenzen – empirische und konzeptionelle Probleme. In: Arbeitsgemeinschaft Qualifikations-Entwicklungs-Management (Hrsg.): Kompetenzentwicklung '99. Aspekte einer neuen Lernkultur, Argumente, Erfahrungen, Konsequenzen, S. 433-493. Münster [u.a.]: Waxmann.

Wild, K.-P. (2000): Lernstrategien im Studium. Münster: Waxmann.

Willenberg, H. (Hrsg.) (2007): Kompetenzhandbuch für den Deutschunterricht. Auf der empirischen Basis des DESI-Projekts. Hohengehren: Schneider.

Wockenfuß, V. & Raatz, U. (2006): Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests. In: R. Grotjahn (Hrsg.): Der C-Test. Theoretische Grundlagen und praktische Anwendungen, S. 211-242. Frankfurt a. M.: Lang.

Wockenfuß, V. (2009): Diagnostik von Sprache und Intelligenz bei Jugendlichen und jungen Erwachsenen. Aachen: Shaker Verlag.

Wollersheim, H.-W. (1993): Kompetenzerziehung : Befähigung zur Bewältigung, Frankfurt am Main [u.a.]: Lang.